

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 28, 2005, 19:17:41 ; Search time 204 Seconds

(without alignments)  
1988.427 Million cell updates/sec

Title: US-09-907-907a-42

Perfect score: 3557

Sequence: 1 DGPFLLPRRRALTOQLVRA.....TAVLHNTQDNERNLILP 705

## Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02:.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3516	98.8	783	2	Q81WX1
2	3512	98.7	783	2	Q8TCS8
3	3280	92.2	783	2	Q8K1R3
4	3033.5	85.3	748	2	Q81R23
5	2783	78.2	647	2	Q8R2U3
6	2389	67.2	540	2	Q9DCS2
7	2285	64.2	504	2	Q96T05
8	1962.5	55.2	784	2	Q7Q228
9	1958	55.0	771	2	Q9V9X7
10	1940.5	54.6	720	2	Q95RX7
11	1761	48.4	720	2	Q95RX7
12	1365.5	38.4	991	2	Q81H29
13	1286.5	36.2	982	2	Q81H29
14	1276	35.9	745	2	Q86SN3
15	1253.5	35.2	707	2	Q8RAK3
16	1212.5	34.1	760	2	Q72ER6
17	1212.5	34.1	760	2	Q72ER6
18	1204.5	33.9	703	2	Q97I45
19	1198	33.7	733	2	Q8KXV3
20	1193.5	33.6	698	2	Q8GKX1
21	1190.5	33.5	698	2	Q99U08
22	1190.5	33.5	698	2	Q99U08
23	1188.5	33.4	698	2	Q7ASX7
24	1188.5	33.4	698	2	Q6G9T9
25	1187.5	33.4	698	2	Q8NMY9
26	1187.5	33.4	713	2	Q9ZAR1
27	1187.5	33.4	713	2	Q7ZJ38
28	1187.5	33.4	713	2	Q7ZJ38
29	1182.5	33.2	722	2	Q6NCN8
30	1182.5	33.2	722	2	Q6NCN8
31	1176	33.1	717	2	Q9ZSW0

32	1175.5	33.0	713	2	Q7NGM0	Q7NGM0
33	1174	33.0	699	2	Q7P6G3	Q7P6G3
34	1172.5	33.0	705	2	Q895J3	Q895J3
35	1172.5	33.0	720	2	Q89WB3	Q89WB3
36	1170	32.9	703	2	Q8RYA1	Q8RYA1
37	1169	32.9	715	2	Q8DGN9	Q8DGN9
38	1167.5	32.8	697	2	Q74CS9	Q74CS9
39	1167.5	32.8	697	2	AAR34967	AAR34967
40	1166.5	32.8	713	2	Q7D283	Q7D283
41	1166.5	32.8	713	2	Q8U56	Q8U56
42	1165.5	32.8	701	2	Q8CST1	Q8CST1
43	1164.5	32.7	697	2	Q6MMS2	Q6MMS2
44	1164.5	32.7	697	2	Q6MMS2	Q6MMS2
45	1163	32.7	717	2	Q8XXP6	Q8XXP6

## ALIGNMENTS

## RESULT 1

ID	Q81WX1	PRELIMINARY:	PRT:	783 AA.
AC	Q81WX1			
DT	01-MAR-2003 (TREMBL)	23, Created		
DT	01-MAR-2003 (TREMBL)	23, Last sequence update		
DT	01-MAR-2004 (TREMBL)	26, Last annotation update		
DE	3'-5' RNA exonuclease.			
CN	Name=OLD35;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed=1247348;			
RA	Leechzynecka M., Kang D.-C., Sarkar D., Su Z.-Z., Holmes M.,			
RA	Valerie K., Fisher P.B.;			
RT	"Identification and cloning of human polynucleotide phosphorylase,			
RT	hnpase (old-35), in the context of terminal differentiation and			
RT	cellular senescence."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16636-16641(2002).			
CC	-1-SIMILARITY: Contains 1 KH domain.			
DR	EMBL; AY027528; AKK13047.1; -.			
DR	HSSP; P05055; ISRO.			
DR	GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.			
DR	GO; GO:0004527; F:exonuclease activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0006396; P:RNA processing; IEA.			
DR	InterPro; IPR001247; 3 Exonase.			
DR	InterPro; IPR001547; Glyco_hydro_5.			
DR	InterPro; IPR004087; KH.			
DR	InterPro; IPR004088; KH_type_1.			
DR	InterPro; IPR008994; Nucleic_acid_DR.			
DR	InterPro; IPR003029; SI.			
DR	Pfam; PF00013; KH 1; 1.			
DR	Pfam; PF03726; RNase; 1.			
DR	Pfam; PF01138; RNase; 2.			
DR	Pfam; PF03725; RNase; PH_C; 2.			
DR	Pfam; PF00575; SI; 1.			
DR	SMART; SM00322; KH; 1.			
DR	SMART; SM00316; SI; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.			
DR	PROSITE; PS50084; KH_TYPE_1; 1.			
DR	PROSITE; PS50126; SI; 1.			
DR	Exonuclease.			
DR	SEQUENCE 783 AA; 85964 MW; 912939AF5309E2C CRC64;			

Query Match 98.8%; Score 3516; DB 2; Length 783;  
Best Local Similarity 99.4%; Pred. No. 1,9e-201;  
Matches 696; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
1 DGPFLLPRRRALTOQLVRAVVDIGNKLEISGKLRFPDGSVAVVGSD 60

```

Db      18 DGFLLPRRRALTOLOVRALMSSAGRAVAVDIGNRKLEISSGKLARFADGSAVVOGSD 77
Qy      61 TAVVAVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
Db      78 TAVVAVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137
Qy      121 IRPLFPAGYFYDYOVLGNLAVDGVNEPDLAINGASVALSLSDIPNMGPGAVRIGIID 180
Db      138 IRPLFPAGYFYDYOVLGNLAVDGVNEPDLAINGASVALSLSDIPNMGPGAVRIGIID 197
Qy      181 GEYVAVNPTRKEMSSSTLNLVAVGAPKSOIWMLEASAEENIIQODFCHAIKVGKVTQOIIQ 240
Db      198 GEYVAVNPTRKEMSSSTLNLVAVGAPKSOIWMLEASAEENIIQODFCHAIKVGKVTQOIIQ 257
Qy      241 GIQOLVKEGTGVTKTPOKLFTPSPEIYKTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db      258 GIQOLVKEGTGVTKTPOKLFTPSPEIYKTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 317
Qy      301 LDTEEOUKERPEADPYEIIIESFNVVAKVFRSIVLNEYKRCDDRDLTSLRNVSCVDYMF 360
Db      318 LDTEEOUKERPEADPYEIIIESFNVVAKVFRSIVLNEYKRCDDRDLTSLRNVSCVDYMF 377
Qy      361 KTLHGSLALFORQGOVLCTVTFDSLESGIKSDOYITAINGIKDKNPMHYEPFYATNEI 420
Db      378 KTLHGSLALFORQGOVLCTVTFDSLESGIKSDOYITAINGIKDKNPMHYEPFYATNEI 437
Qy      421 GKVTGLNRRELGHGALAEKALPYIPRDPFTTIRVTSEVLESNGSSMASACGSLAMD 480
Db      438 GKVTGLNRRELGHGALAEKALPYIPRDPFTTIRVTSEVLESNGSSMASACGSLAMD 497
Qy      481 SGVIVSSAVAGVALGYTKTDPKEGGEIEDRYLLDIIIGIEBYNDMPDKAGTKKGTAL 540
Db      498 SGVIVSSAVAGVALGYTKTDPKEGGEIEDRYLLDIIIGIEBYNDMPDKAGTKKGTAL 557
Qy      541 QADIKLKGIPKIKYWEAIIQASVAKKEILQIMNKTIKSPRASRKENGVPVETVOVPIKSR 600
Db      558 QADIKLKGIPKIKYWEAIIQASVAKKEILQIMNKTIKSPRASRKENGVPVETVOVPIKSR 617
Qy      601 AKFVPGGYNLKKLQIAGTGVTTISQVDEBTFSVFAPTSPVMEHAPDTTEICKDQEOOLE 660
Db      618 AKFVPGGYNLKKLQIAGTGVTTISQVDEBTFSVFAPTSPVMEHAPDTTEICKDQEOOLE 677
Qy      661 FGAVYVATITIRPTGVWVKIYPMNTAVLNLNTOLDNERL 700
Db      678 FGAVYVATITIRPTGVWVKIYPMNTAVLNLNTOLDNERL 717

RESULT 2
ID      08TCS8      PRELIMINARY;      PRT;      783 AA.
AC      08TCS8;
DT      01-JUN-2002 (TREMBlrel. 21, Created)
DT      01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT      01-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE      Polynucleotide phosphatylase-like protein (EC 2.7.7.8)
DE      (Polynucleotide nucleotidyltransferase 1).
GN      Name=PNPASE; Synonyms=PNP1;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      MEDLINE=22307406; Pubmed=12419256;
RA      Rajmehers R., Vree Esberts W., van Venrooij W., Pruijn G.;
RT      "Protein-protein interactions between human exosome components support
RT      the assembly of RNAse PH-type subunits into a six-membered PNPase-like
RT      ring."
RL      J. Mol. Biol. 323:653-663 (2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RX      MEDLINE=22386257; Pubmed=12477932;

```

```

RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marisita K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udén T.B., Toehiyuki S., Carninci P., Prange C.T.,
RA      Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA      Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences."
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RA      Strausberg R.;
RL      Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC      -1- Similarity: Contains 1 KH domain.
DR      EMBL; AJ458465; CAD30289.1; -.
DR      EMBL; BC053660; AAH53660.1; -.
DR      PIR; T50626; T50626.
DR      HSSP; P05055; ISRO.
DR      GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR      GO; GO:0004654; F:polynucleotide nucleotidyltransferase a. .; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006396; P:RNA processing; IEA.
DR      InterPro; IPR001247; 3 Exonase.
DR      InterPro; IPR001547; Glyco_hydro_5.
DR      InterPro; IPR004087; KH.
DR      InterPro; IPR004088; KH_type_1.
DR      InterPro; IPR008994; Nucleic_acid_ob.
DR      InterPro; IPR003029; S1.
DR      Pfam; PF00013; KH 1; 1.
DR      Pfam; PF03726; PNPase; 1.
DR      Pfam; PF01138; RNase_PH; 2.
DR      Pfam; PF03725; RNase_PH_C; 2.
DR      Pfam; PF00575; S1; 1.
DR      SMART; SM00322; KH; 1.
DR      SMART; SM00316; S1; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR      PROSITE; PS50084; KH_TYPE_1; 1.
DR      PROSITE; PS50126; S1; 1.
KW      Nucleotidyltransferase; Transferase.
SQ      SEQUENCE 783 AA; 85936 MW; 8A3629AF52F8E24 CRC64;

Query Match      98.7%; Score 3512; DB 2; Length 783;
Best Local Similarity 99.3%; Pred. No. 3.4e-201;
Matches 695; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 241 GIGQVKEGVTKTKTPKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300  
 DB 258 GIGQVKEGVTKTKTPKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317  
 QY 301 LDTBEOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCDSGRDLTSLRNVSCVDMP 360  
 DB 318 LDTBEOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCDSGRDLTSLRNVSCVDMP 377  
 QY 361 KTLHGSALFORGOVCTVTPDSLESIGSDQVITANGIKDNPMFLHYEPPYATNEI 420  
 DB 378 KTLHGSALFORGOVCTVTPDSLESIGSDQVITANGIKDNPMFLHYEPPYATNEI 437  
 QY 421 GKVTGLRREIGHALAEKALPYIPDPFPTITVTSVLESNGSSMASACGSLALMD 480  
 DB 438 GKVTGLRREIGHALAEKALPYIPDPFPTITVTSVLESNGSSMASACGSLALMD 497  
 QY 481 SGVPISAVAGVALGVTCTDPEKGEIEDYRLTDIGIEDYNGDMDFKTAGNKGITLAL 540  
 DB 498 SGVPISAVAGVALGVTCTDPEKGEIEDYRLTDIGIEDYNGDMDFKTAGNKGITLAL 557  
 QY 541 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTTISKPRASRENGPVVETVQVPLSKR 600  
 DB 558 QADIKLPGIPKIYMEAIQOASVAKKEILOIMNTTISKPRASRENGPVVETVQVPLSKR 617  
 QY 601 AKFPVPGGVNKKLQAEFTGVTISQVDEFTSFVAPTPSVMEHARDFTIEICKDOEOOLE 660  
 DB 618 AKFPVPGGVNKKLQAEFTGVTISQVDEFTSFVAPTPSVMEHARDFTIEICKDOEOOLE 677  
 QY 661 FGAVYATITEIRDTGVVVKLPMNTAVLHNTOLDNERL 700  
 DB 678 FGAVYATITEIRDTGVVVKLPMNTAVLHNTOLDNERL 717

## RESULT 3

Q8KIR3 PRELIMINARY; PRT; 783 AA.  
 ID Q8KIR3  
 AC Q8KIR3  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Polynucleotide phosphorylase (EC 2.7.7.8).  
 GN Name=ppp1; Synonyms=PNPase;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=divers;  
 RX MEDLINE=22307406; PubMed=12419256;  
 RA Rajmehkerts R., Vree Egberts W., van Venrooij W., Pruijn G.;  
 RT "Protein-protein interactions between human exosome components support  
 the assembly of RNAse PH-type subunits into a six-membered PNPase-like  
 ring";  
 RT J. Mol. Biol. 323:653-663(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheiner C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stadelton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abraham R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Rodríguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,  
 RA Krzywnicki M.I., Skalka U., Smalms D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RA Strausberg R.;  
 RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- Similarity: Contains 1 KH domain.  
 DR EMBL; AJ507387; CAD5436.1; -;  
 DR EMBL; BC055826; AAH55826.1; -;  
 DR HSSP; P05055; ISRO.  
 DR MOD; MGI.1918951; Pnp1.  
 DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.  
 DR GO; GO:0004654; F:Polyribonucleotide nucleotidyltransferase a. . .; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006396; P:RNA processing; IEA.  
 DR InterPro; IPR001247; 3-ExoRNase.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR InterPro; IPR004087; KH.  
 DR InterPro; IPR004088; KH\_type\_1.  
 DR InterPro; IPR008994; Nucleic\_acid\_OB.  
 DR InterPro; IPR003029; S1.  
 DR Pfam; PF00013; KH\_1; 1.  
 DR Pfam; PF03726; PNPase; 1.  
 DR Pfam; PF0138; RNase\_PH; 2.  
 DR Pfam; PF03725; RNase\_PH\_C; 2.  
 DR Pfam; PF00575; S1; 1.  
 DR SMART; SM00322; KH; 1.  
 DR SMART; SM00316; S1; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN\_1.  
 DR PROSITE; PS50084; KH\_TYPE\_1; 1.  
 DR PROSITE; PS50126; S1; 1.  
 KW Nucleotidyltransferase; Transferase.  
 SQ SEQUENCE 783 AA; 85682 MW; F35FEBE91AAB5626 CRC64;

Query Match 92.28; Score 3280; DB 2; Length 783;  
 Best local similarity 91.08; Pred. No. 2.5e-187;  
 Matches 636; Conservative 39; Mismatches 24; Indels 0; Gaps 0;

QY 2 GPTLPDRDALTOLOVRALMSSAGSAVAVVDIGNRLKLEISSGKLAPADGCAVVGSDT 61  
 DB 19 GPTLPDRDALTOLOVRALMSSAGSAVAVVDIGNRLKLEISSGKLAPADGCAVVGSDT 78  
 QY 62 AVMTAVSKTKPSPQMPVVDYRQKAAAGRIPTYLRREVSTDSKEILTSRIIDRSI 121  
 DB 79 AVMTAVSKTKKASQMPVVDYRQKAAAGRIPTYLRREVSTDSKEILTSRIIDRSI 138  
 QY 122 RPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLDSIPNNGPVAGVRIIGD 181  
 DB 139 RPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLDSIPNNGPVAGVRIIGD 198  
 QY 182 EYVNPTRKEMSSSTLNLVVAAGPKQIWMLESAENIILQDPCHAIKGVTKYQOIIIG 241  
 DB 199 ECVNPTRRREMSSTLNLVVAAGPKQIWMLESAENIILQDPCHAIKGVTKYQOIIIG 258  
 QY 242 IQQLVKETGYTKRPQKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 301  
 DB 259 IQQLVKETGYTKRPQKLTFTSPSEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 318  
 QY 302 DTEBOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCDSGRDLTSLRNVSCVDMP 361  
 DB 319 DTEBOLKEKPEADPEYIEISFNVAKEVFRSIVLNEYKRCDSGRDLTSLRNVSCVDMP 378  
 QY 362 TLHGSALFORGOVCTVTPDSLESIGSDQVITANGIKDNPMFLHYEPPYATNEIG 421  
 DB 379 TLHGSALFORGOVCTVTPDSLESIGSDQVITANGIKDNPMFLHYEPPYATNEIG 438

QY 422 KVTGLNRELGHGALAEKALYPIVPRDPPTIRTVTSVLEBSNGSSSMASACGSLALMDS 481  
 DB 439 KVTGVNRRRELGHGALAEKALCPVLPKDFPTIRTVTSVLEBSNGSSSMASACGSLALMDS 498  
 QY 482 GVPISAVAGVAGLVTITDPEKEIEDYRLITDILGIEDYNGMDPKIAGTNGITATLQ 541  
 DB 499 GVPISAVAGVAGVAVLVTITDPEKEIEDYRLITDILGIEDYNGMDPKIAGTNGITATLQ 558  
 QY 542 ADIKLPGLPIKIVMEALIQOASVAKKEILQIMNKTIKPRASRKENGVEVTVQVPLSKRA 601  
 DB 559 ADIKLPGLPIKIVMEALIQOASVAKKEILQIMNKTIKPRASRKENGVEVTVQVPLSKRA 618  
 QY 602 KFPVPGGVNKKLQAEFGVITISQVDEETFSFAPTPSPVMEHARDFITEICKDQEQLEF 661  
 DB 619 KFPVPGGVNKKLQAEFGVITISQVDEETFSFAPTPSPVMEHARDFITEICKDQEQLEF 678  
 QY 662 GAVYTATITEIRDTGVNWKLYPNMTAVLHNTQIDNRL 700  
 DB 679 GAVYTATITEIRDTGVNWKLYPNMTAVLHNTQIDNRL 717

## RESULT 4

0812B3 ID 0812B3 PRELIMINARY; PRT: 748 AA.

AC 0812B3; 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Polynucleotide phosphorylase.  
 GN Name=Pnpl1; Synonyms=Oid35.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Leecezyniecka M., Fisher P.B.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 KH domain.  
 DR HSSP; P05055; ISRO.  
 DR MGD; MGI:1918951; Pnpl1.  
 DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0006396; P:RNA processing; IEA.  
 DR InterPro; IPR001247; 3 ExorNae.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR InterPro; IPR004087; KH\_type\_1.  
 DR InterPro; IPR004088; KH\_type\_1.  
 DR InterPro; IPR008994; Nucleic\_acid\_OB.  
 DR InterPro; IPR003029; S1.  
 DR Pfam; PF00013; KH\_1; 1.  
 DR Pfam; PF01118; RNasec\_P1; 2.  
 DR Pfam; PF03725; RNasec\_P1; 2.  
 DR Pfam; PF00575; S1; 1.  
 DR SMART; SM00322; KH; 1.  
 DR SMART; SM00316; S1; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN\_1.  
 DR PROSITE; PS50084; KH\_TYPE\_1; 1.  
 DR PROSITE; PS50126; S1; 1.  
 SQ SEQUENCE 748 AA; 81959 MW; 4D82B64C5DBBE769 CRC64;

Query March 85.3%; Score 3033.5; DB 2; Length 748;  
 Best Local Similarity 87.0%; Pred. No. 1.2e-172;  
 Matches 596; Conservative 41; Mismatches 43; Indels 5; Gaps 2;  
 QY 18 VRAIWSAGSRAVAVDIGNRLKELISSGKLAFADGSAVVOGDTAVAVTAVSKTSPSPQ 77  
 DB 1 MRALMSSTGRRAVAVDIGNRLKELISSGKLAFADGSAVVOGDTAVAVTAVSKTSPSPQ 60  
 QY 78 FMPVIVYVYRQKAAAGRIPTNYIRREVGTSKKEILTSRIIDRSIRLPFPAGYFDTQVLC 137  
 DB 61 FMPVIVYVYRQKAAAGRIPTNYIRREVGTSKKEILTSRIIDRSIRLPFPAGYFDTQVLC 120

QY 138 NLAVNGVNEPDLAINAGASVALSLSDIPNNGVGVAVRIGIIDGEVYVNPTRKEMSSTL 197  
 DB 121 NLAVNGVNEPDLAINAGASVALSLSDIPNNGVGVAVRIGIIDGEVYVNPTRKEMSSTL 180  
 QY 198 NLVAVGAPKSIQVWLFRASVENILQODPCHAIKGVVYTOOIIQGIQOLVETVTKRTPQ 257  
 DB 181 NLVAVGAPKSIQVWLFRASVENILQODPCHAIKGVVYTOOIIQGIQOLVETVTKRTPQ 240  
 QY 258 KLPSPSEIVKYTHKLAMERLYAVFTDYEDHKVSRDEAVN--KIRLDTEQLKEKPEAD 315  
 DB 241 EDIYSPFCRDEYVTKXIAMETLYAGFQIMMIXF---GNNCYGIRLDPRIYKEKPEVD 297  
 QY 316 PYTIISEFNVAKEVRSIYLANEYKQCDGRLTSLKRVSCEDUMFTLHSAALPQKQTO 375  
 DB 298 QPEIIIESFNIVAKEVRSIILNEYKQCDGRLTSLKRVSCEDUMFTLHSAALPQKQTO 357  
 QY 376 VLCTVFPDSLESGIKSDQVLTAINGIKDKNFMHYEPFPATNEIGKVGYNRELGHGA 435  
 DB 358 VLCTVFPDSLESGIKSDQVLTAINGVADKNFMHYEPFPATNEIGKVGYNRELGHGA 417  
 QY 436 LAEKALYPIVPRDPPTIRTVTSVLEBSNGSSSMASACGSLALMDSGVPISSAVAGVAG 495  
 DB 418 LAEKALCPVLPKDFPTIRTVTSVLEBSNGSSSMASACGSLALMDSGVPISSAVAGVAG 477  
 QY 496 LVTKTDPKEKEIEDYRLITDILGIEDYNGMDPKIAGTNGITATLQADIKLPGLPIKIVM 555  
 DB 478 LVTKTDPKEKEIEDYRLITDILGIEDYNGMDPKIAGTNGITATLQADIKLPGLPIKIVM 537  
 QY 556 EATQOASVAKKEILQIMNKTIKPRASRKENGVEVTVQVPLSKRAKFPVGGVYNKKIQ 615  
 DB 538 EATQOASVAKKEILQIMNKTIKPRASRKENGVEVTVQVPLSKRAKFPVGGVYNKKIQ 557  
 QY 616 AETGVITISQVDEETFSFAPTPSPVMEHARDFITEICKDQEQLEFGAVYTATITEIRDT 675  
 DB 598 AETGVITISQVDEETFSFAPTPSPVMEHARDFITEICKDQEQLEFGAVYTATITEIRDT 657  
 QY 676 GAVYTATITEIRDTGVNWKLYPNMTAVLHNTQIDNRL 700  
 DB 658 GAVYTATITEIRDTGVNWKLYPNMTAVLHNTQIDNRL 682

## RESULT 5

08R2U3 ID 08R2U3 PRELIMINARY; PRT: 647 AA.

AC 08R2U3; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Pnpl1 protein (Fragment).  
 GN Name=Pnpl1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=FVB/N;  
 RC TISSUE=Mammary tumor, Metastathionien-TGF alpha model. 10 month old  
 RC virgin mouse, taken by biopsy;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins T.B., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueidi T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fehey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzyzanski M.I., Skalska U., Smalley D.E., Scherch A., Schain J.E.,  
 RA Jones S.U., Warr M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N;  
 RC TISSUE=Mammary tumor. Metallochromen-TGF alpha model. 10 month old  
 RC virgin mouse. Taken by biopsy.;  
 RA Strauberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 KH domain.  
 DR EMBL: BC027228; AAR27228.2; -.  
 DR MGI: MGI:1918951; Pppl1.  
 DR GO: GO:0000175; F:3'-5'-exoribonuclease activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0006396; P:RNA processing; IEA.  
 DR InterPro: IPR001247; 3-Exonase.  
 DR InterPro: IPR001547; Glyco\_hydro\_5.  
 DR InterPro: IPR004088; KH\_type\_1.  
 DR InterPro: IPR008994; Nucleic\_acid\_OB.  
 DR InterPro: IPR003029; S1.  
 DR Pfam: PF00013; KH\_1; 1.  
 DR Pfam: PF03726; NPase; 1.  
 DR Pfam: PF01138; RNase\_PH; 2.  
 DR Pfam: PF03725; RNase\_PH\_C; 2.  
 DR Pfam: PF00575; S1; 1.  
 DR SMART: SM00332; KH; 1.  
 DR SMART: SM00316; S1; 1.  
 DR PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN\_1.  
 DR PROSITE: PS50084; KH\_TYPE\_1; 1.  
 DR PROSITE: PS50126; S1; 1.  
 DR NON TER 1  
 SQ SEQUENCE 647 AA; 71015 MW; 5B1574DBDA723B43 CRC64;

Query Match 78.2%; Score 2783; DB 2; Length 647;  
 Best Local Similarity 92.4%; Pred. No. 9.2e-156;  
 Matches 537; Conservative 29; Mismatches 15; Indels 0; Gaps 0;

QY 120 SIRLPAGYPTDQVLCNLAVDGVNEPDLAINGSVALSLSDIPMGVGVARRIGI 179  
 DB 1 SIRLPAGYPTDQVLCNLAVDGVNEPDLAINGSVALSLSDIPMGVGVARRIGI 60  
 QY 180 DGEVVPTRKEMSSSTLNLVAVAGPKSQIWMLEASANIQQDFCHAIKGVKYYTQOI 239  
 DB 61 DGEVVPTRKEMSSSTLNLVAVAGPKSQIWMLEASANIQQDFCHAIKGVKYYTQOI 120  
 QY 240 QGIQQLVKEGTGVTTRTPQKLFPTSPETVYKTHKLAMERLYAVFTDYEDHKVSRDEAVNKI 299  
 DB 121 QGIQQLVKEGTGVTTRTPQKLFPTSPETVYKTHKLAMERLYAVFTDYEDHKVSRDEAVNKI 180  
 QY 300 RLDTREOLKEKFPADPEIETESFNVAKEVFRSIVLNEYKRCGGRDLTSLRANSCEDVM 359  
 DB 181 RLDTREOLKEKFPADPEIETESFNVAKEVFRSIVLNEYKRCGGRDLTSLRANSCEDVM 240  
 QY 360 FKTLHSGALFORGTOVLCVTPDLSLESGISDQVITANGIKDKNFMLATPEPPYATNE 419  
 DB 241 FKTLHSGALFORGTOVLCVTPDLSLESGISDQVITANGIKDKNFMLATPEPPYATNE 300  
 QY 420 IGVKTGNRLRELGALAEKALYVPIPRDPFTIRVTSVLESNGSSMASACGSLALM 479  
 DB 301 TGVKTGNRLRELGALAEKALYVPIPRDPFTIRVTSVLESNGSSMASACGSLALM 360  
 QY 480 DSGVPISSAVAGVAILGVTKDPEKGEIEDRLTDLIGIDNNGDMFKTAGNKGTA 539  
 DB 361 DSGVPISSAVAGVAILGVTKDPEKGEIEDRLTDLIGIDNNGDMFKTAGNKGTA 420  
 QY 540 LQADIKLPGVIFIKIWEAIIQASVAKKEIIQIMNKTISKPRASREKNGPVETQVPLSK 599

DB 421 LQADIKLPGVIFIKIWEAIIQASVAKKEIIQIMNKTISKPRASREKNGPVETQVPLSK 480  
 QY 600 RAKFVGGVNLKQLQATGTTISQVDEEFPSSVAPPSVWEHARDFTETCXDDOQOL 659  
 DB 481 RAKFVGGVNLKQLQATGTTISQVDEEFPSSVAPPSVWEHARDFTETCXDDOQOL 540  
 QY 660 EFGAVYATITTEIRDTCVMVLYPNMTAVLLHNTQNLNERL 700  
 DB 541 EFGAVYATITTEIRDTCVMVLYPNMTAVLLHNTQNLNERL 581

RESULT 6  
 Q9DC52 PRELIMINARY; PRT; 540 AA.  
 AC Q9DC52.  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus adult male lung cDNA, RIKEN full-length enriched library,  
 DE clone:120003F12 product:hypothetical 3' exoribonuclease family  
 DE containing protein, full insert sequence.  
 GN Name=Proct;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RA The FANTOM Consortium;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Lung;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Mizawa K., Nagoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanabe M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-formet  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Lung;
RA Aachai J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuo M.,
RA Hanagaki T., Hara A., Harauchi N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishi Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai U., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sojabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK004563; BAB3374.1; -.
DR HSSP; Q53597; IE3P.
DR MGD; MGI:1918951; Pnpf1.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3 ExoRNase.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KM HYPOthetical protein.
SQ SEQUENCE 540 AA; 58938 MW; 457BFA3E3579A072 CRC64;

Query Match 67.2%; Score 2389; DB 2; Length 540;
Best Local Similarity 90.1%; Pred. No. 2.5e-134;
Matches 465; Conservative 29; Mismatches 22; Indels 0; Gaps 0;

QY 2 GPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSGD 61
DB 19 GFLRPGGRNRALSTLOVRALWSSGSRVAVYDLDHRLKLEISSGKLARFADGSAVVGSGD 78
QY 62 AAVTAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSI 121
DB 79 AAVTAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRSI 138
QY 122 RPLFPAGFYDTQVLCNLLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIIDG 181
DB 139 RPLFPAGFYDTQVLCNLLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIIDG 198
QY 182 EYVNPTRKEMSSSTLNLVVGAKPSQIVMLEASAEINLODPCHAIKVGKTYQOIIQG 241
DB 199 ECVNPFRRKEMSSSTLNLVVGAKPSQIVMLEASAEINLODPCHAIKVGKTYQOIIQG 258
QY 242 IOQLVKEGTGVTTPQKLFTPSPEIIVKXTHKLAMERLVAFTDYEHDKSVSDAANKIRL 301
DB 259 IOQLVKEGTGVTTPQKLFTPSPEIIVKXTHKLAMEKLVAFTDYEHDKSVSDAANKIRL 318
QY 302 DTEQLKKEPPADPYEIIIESFNVAKEVPSIYLNRYKRDGDLTSLRVNCEVDMEK 361
DB 319 DTEHLKKEPPADPYEIIIESFNVAKEVPSIYLNRYKRDGDLTSLRVNCEVDMEK 378
QY 362 TLHGSAIFORGQOVLCVTFPSLSEGIKSDQVITAINGIKDKPMHLHYEPPTATNEIG 421
DB 379 TLHGSAIFORGQOVLCVTFPSLSEGIKSDQVITAINGIKDKPMHLHYEPPTATNEIG 438
QY 422 KVTGLNRELGHGALAEKALVYIPDPFPTIRVTSVELESNGSSVMAACGGSIALMDS 481
DB 439 KVTGLNRELGHGALAEKALVYIPDPFPTIRVTSVELESNGSSVMAACGGSIALMDS 498
QY 482 GVPISSAVAGVAGLVTKTDEKGIIDYRLITDIL 517
DB 499 GVPISSAVAGVAGLVTKTDEKGIIDYRLITDIL 534

```

```

AC Q96T05;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
DE Hypothetical protein FLJ14531.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI_TaxID=9606;
RX PubMed=14702039;
RP SEQUENCE FROM N.A.
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiyama S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Miasashiro K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Sasaki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukumizu Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujitani T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohnori Y.,
RA Kawabata A., Hikiji T., Kobayashi N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtsuka R., Kawakami T., Noguchi S., Itoh T., Shigemura K., Sena T.,
RA Matsunaga K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK027437; BAB55109.1; -.
DR HSSP; Q53597; IE3H.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3 ExoRNase.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KM transferase.
SQ SEQUENCE 504 AA; 55998 MW; 2BB89ADB409322D6 CRC64;

Query Match 64.2%; Score 2285; DB 2; Length 504;
Best Local Similarity 100.0%; Pred. No. 3.7e-128;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSGD 60
DB 18 DGPFLLPRDRALTOLOVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVGSGD 77
QY 61 TAAVNTAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRS 120
DB 78 TAAVNTAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVSTDEKILTSRIIDRS 137
QY 121 IRPLFPAGFYDTQVLCNLLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
DB 138 IRPLFPAGFYDTQVLCNLLAVDGVNEPVLAINGASVALSLSDIPMNGPVGAVRIGIID 197

```

```

Oy 181 GEYVNFTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 240
Db 198 GEYVNFTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 257
Oy 241 GIGQVKEGTGVTTRPQKLTFTSPBEIVKTYTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db 258 GIGQVKEGTGVTTRPQKLTFTSPBEIVKTYTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 317
Oy 301 LDTEBOLKEKPPADPEIIESFNVAKEVRSIVLNEYKRCDDGDLTSLRNVSCEDVDF 360
Db 318 LDTEBOLKEKPPADPEIIESFNVAKEVRSIVLNEYKRCDDGDLTSLRNVSCEDVDF 377
Oy 361 KTLHGSALFORGOVCTVTFPSLSGSKSDOYITAINIKDKNFMPLHYEPFYATNEI 420
Db 378 KTLHGSALFORGOVCTVTFPSLSGSKSDOYITAINIKDKNFMPLHYEPFYATNEI 437
Oy 421 GKVTGLNRRELGHGALAEKALYPVIRPDPF 450
Db 438 GKVTGLNRRELGHGALAEKALYPVIRPDPF 467

RESULT 8
ID 070228 PRELIMINARY; PRT; 784 AA.
AC 070228;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE AGCP9005.
GN Name=agCG49269; ORFNames=ENSAG00000018771;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
   EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
   preliminary data.
CC EMBL: AAB01008979; EMBL31718.1;
DR GO: GO:0000175; F:3'-5'-exonibonuclease activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0006396; P:RNA processing; IEA.
DR InterPro: IPR001247; 3 ExonName.
DR InterPro: IPR004088; KH_type_1.
DR InterPro: IPR003029; SL.
DR Pfam: PF00013; KH_1.1.
DR Pfam: PF03726; ENase_1.
DR Pfam: PF01138; RNase_PH_2.
DR Pfam: PF03725; RNase_PH_2.
DR PROSITE: PSS0084; KH_type_1; 1.
DR PROSITE: PSS0125; SL_1.
SO SEQUENCE 784 AA; 86050 MW; D09241BF3A1C4A30 CRC64;

Query Match 55.2%; Score 1962.5; DB 2; Length 784;
Best Local Similarity 55.4%; Pred. No. 1.3e-108;
Matches 387; Conservative 119; Mismatches 187; Indels 5; Gaps 4;
Oy 5 LLPR-RBALTOLOVALMSAGRAVAVDLGNRLKLEISSGKLARPADGSVAVOSGDTAV 63
Db 18 ILPHVRCATRRQSTSLSPSEVD--VALSTG-RILKISSGKVARFADGCSVTIGDTAV 74
Oy 64 MVTAVSTKPSQFMPLVVDYRQKAAAGRIPTNYLARBVGSTDKELTIRSIIDRSIRP 123
Db 75 MVTAVAKOKSNASFLVVDYRQKAAAGRIPTNYLARBVGSTDKELTIRSIIDRSIRP 134
Oy 124 LFPAGVYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNMGVAVRIGIIDGEX 183
Db 135 LFPABFYDTQVLCNMLAIDGAPPDVQAINGASVALSLSDIPNMGVAVRIGIIDGEX 194

```

```

Oy 184 VVNFTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 243
Db 195 IINPTRKEMSSSTLNLYVAGAPKSOIYMLBASAEINILOQDFCHAIKGVKXTQOIIQ 254
Oy 244 QLVKENGVTGVTTRPQKLTFTSPBEIVKTYTHKLMERLYAVFTDYEHDKVSRDEAVNKIR 303
Db 255 RLQGVKPKRRLAEPPLAVDGEIIONAVQTSSEMRRLREIFDYSHDKPSRDOAVSKTRTDT 314
Oy 304 BEOLAKEPPADPEIIESFNVAKEVRSIVLNEYKRCDDGDLTSLRNVSCEDVDF 360
Db 315 IDKWSSFPADPGVISTETNKVKVGFREMLLEKRCRGGDLDRKISCVNLHKL 374
Oy 364 HGSALFORGOVCTVTFPSLSGSKSDOYITAINIKDKNFMPLHYEPFYATNEI 423
Db 375 HGSALFORGOVCTVTFPSLSGSKSDOYITAINIKDKNFMPLHYEPFYATNEI 434
Oy 424 TGLNRRELGHGALAEKALYPVIRPDPFTIRVTSVLNSGSSMSACGSLALDSGV 483
Db 435 GPGRIRRELGHGALAEKALYPVIRPDPFTIRVTSVLNSGSSMSACGSLALDSGV 494
Oy 484 PISAVAGVAILGVTR-TDEKGBIEDYRLTLGIEDYNGMDPFIAGTKGITLQA 542
Db 495 PVQEAAGVAILGVTR-TDEKGBIEDYRLTLGIEDYNGMDPFIAGTKGITLQA 554
Oy 543 DIKLPGRIPKIMVIAIQASVAKKEILQIMNKITSKPRASKENGPVETVOVPLSRKAK 602
Db 555 DIKLPGRIPKIMVIAIQASVAKKEILQIMNKITSKPRASKENGPVETVOVPLSRKAK 614
Oy 603 FVPGGYNLKRLQAEVTGTSOVDERTFVAPPTPSVMEARDFTEICDDOEOLEFG 662
Db 615 LIFGGGTNLRLVLEGTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVT 674
Oy 663 AVTATITTEIRDTGVKYLKYPMTAVILNTQDLNREL 700
Db 675 ALTYARLVELDGTGVWTLVPSMPTLHNSQDLQKRI 712

RESULT 9
ID 09V9X7 PRELIMINARY; PRT; 771 AA.
AC 09V9X7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG11337-PA (CG11337-pb).
GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; Pubmed=10711132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Baas J., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA DePablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glaeser K.,
RA Foster C., Gabriellian A.E., Garcia N.S., Gelbart W.M., Glaeser K.,
RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harrie N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```



RA Hostin K., Houston F.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jatali M., Kalush K., Kapen G.H., Ke Z., Kennison J.B., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Liao X., Lei Y., Levitsky A.A., Li Y., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Mishina N.V., Moberly A.C., Morris J., Moshier A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Patel J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.W., Woodager, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "the genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2242605; PubMed=12537568;  
RA Ceinkner S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,  
RA George K.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
RA Patel J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Swirskas R., Taor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT *melanogaster* euchromatic genome sequence";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminer J.S., Bergman C.M., Krommiller B., Carlson J., Swirskas R.,  
RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Ceinkner S.E.;  
RT The transposable elements of the *Drosophila melanogaster* euchromatin:  
RL a genomic perspective.";  
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2242609; PubMed=12537572;  
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminer J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu I., Berman B.P.,  
RA Bettencourt B.R., Ceinkner S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
[5]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
CC Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
-1- SIMILARITY: Contains 1 KH domain.  
DR EMBL; AE003778; AAF57151.2; -.  
DR HSSP; Q53597; IE3P.  
DR InFACt; Q9V9X7; -.  
DR FLYBase; FBgn0039846; CG11337.  
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0006396; P:RNA processing; IEA.  
DR InterPro; IPR001247; 3\_ExonBase.  
DR InterPro; IPR004087; KH.  
DR InterPro; IPR004088; KH type 1.

DR	InterPro; IPR006994; Nucleic_acid_08.
DR	InterPro; IPR003029; SL.
DR	Pfam; PF00013; KH_1. 1.
DR	Pfam; PF03726; ENPase; 1.
DR	Pfam; PF01138; RNase_PH; 2.
DR	Pfam; PF03725; RNase_PH_C; 2.
DR	SMART; SMO0322; KH; 1.
DR	SMART; SMO0326; SL; 1.
DR	PROSITE; PS50084; KH_type_1; 1.
DR	PROSITE; PS50126; SL; 1.
DR	SEQUENCE 771 AA; 84953 MW; 779F086DB17022C4 CRC64;
Query Match	55.0%; Score 1958; DB 2; Length 771;
Best Local Similarity	55.2%; Pred. No.2.4e-108;
Matches 380; Conservative 128; Mismatches 173; Indels 8; Gaps 5	
OY	19 RALWSAGSRAVAVDL--GNRKLEISSGKLARFADGSAAVVGSDTAVMTVAVSXTKSP 75
Dd	29 RGIGSSSGAEPASVEAVNFSNGRNTPFSSGRLARAFANGTAVCQMGDTAVMTVAKAKNP 88
OY	76 SQ-FMPLYVDYRQAAAAGRIPNYLRREVGTSDKEILTSTRIIDRSIRPLFPAGFYDTQ 134
Dd	89 GGCEMPLVDYRLKAASGRIPMFMRRELTPSEKEILISARLIRSLPLPFKDRTREQ 148
OY	135 VLCLNLAVDGNEDDYLAINGASVALSDI.PMWGPVGAIVIGITDGCVVNPTKEMSS 194
Dd	149 LVCCNLADAHAHSPVLIALINAPSAHLSDI.PWNGPIDAVRGLCDGEVLINFTIRELOT 208
OY	195 STLNIVVAGAASKSOIVMLEASAENILODFCHAIKVGVKYTOOIQIOQLVKENGVTKR 254
Dd	209 SQLDLVVASATQNLVLMLEGKNVLOQDDLKAIKGGRBAQFIHEIERLOKAVGRQR 268
OY	255 TPQKLFPPSPPIVKYTKHAMERLYAVFTDYEHDKVSDFAVNRKRLDTEOLKKFFEA 314
Dd	269 EEEVAAEVDPELGKAVASMCMERLEIFQDSTHDMSRDNAVNEVRSHVIDKWSSFPDT 328
OY	315 DPEYLIESFNVAKEVPRIYLVNEYKRGDDGLDSLNVSCGEVMFKTLHSALFORQOT 374
Dd	329 EPSDITEQNFQTSRTIFRELLFERGLRCDDGRDYDQLRNISCVDMYKPLHGSALEFORQT 388
OY	375 QVLCTVFPSDESIGSKPOVITAIN-GIKQKNFMIAHPYPYATNEIGKTVGLNRBELG 432
Dd	389 QVCFCTVISLDSBSAMKIDS-LAALDSGLKAKNPMFLHIEFPFYANGEVGRIGPVRREMGG 447
OY	433 HGALAERKALYFVIPRDPFFTIRVTSEVLESNGSSSMASACGGSIALMDGVPISAAVAGV 492
Dd	448 HGALEARSLPLTLPNDYEFYTRLNSEVLESNGSSMASVCGSIALMDAGVAPSAPAAGV 507
OY	493 AIGLVTK-TDEKGEIEDYRLTLDIGIEDNGMDFLAGINKKITLQNDIKLPGPI 551
Dd	508 AIGLVTKENDTDKHLQDYRLITLDILGIEDYGMDDMKVAGRKFTAIQAALKLPGPL 567
OY	552 KIWEAIDQASVAKKEILQIMNKITSKPASRKENGPVETQVPLSRKRVFGGYNL 611
Dd	568 KVMESLDAKADANSNIIDIMSAILREKRKPKESWPASELITVBPQQBAGLIGSGLHM 627
OY	612 KKLQALETGVTISOYDEERTFVPAPTPSVMEARDPITEIKCDDOEQOLEFGAVTYATITE 671
Dd	628 KRIVLETGSTLAYDETHTFNVPAPQAMDCAKEKELIBGVYWKERVBDLEFGGITRYAKITE 687
OY	672 IRDTGVWVKLYPNMTAVLIANTOLDNELU 700
Dd	688 LRDTGVWVILYPSMPALLHNSOLDORCI 716
RESULT 10	
G9SRX7	PRELIMINARY; PRT; 720 AA.
AC	G9SRX7;
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	LDD3255P.



```

GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
NCBI_Locus=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorese V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,
RA Munoz J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Cejnkner S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 KH domain.
CC EMBL; AY061061; AAL28609.1; -.
DR HSSP; Q53597; 1E3P.
DR FlyBase; FBgn0039846; CG11337.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3_ExonNase.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH type 1.
DR InterPro; IPR008994; Nucleic_acid_ob.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH_1; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PSS0084; KH_TYPE_1; 1.
DR PROSITE; PSS0126; S1; 1.
DR PROSITE; PSS0126; S1; 1.
SQ SEQUENCE 720 AA; 79381 MW; DD54952528ESDBDF CRC64;

Query March 54.6%; Score 1940.5; DB 2; Length 720;
Best Local Similarity 56.2%; Pred. No. 2.4e-107;
Matches 374; Conservative 123; Mismatches 164; Indels 5; Gaps 4;

QY 39 LEISSGLKARPADSSAVVQSGDTAVMTAVSKTRPSBQ-EMPIVVDYRQKAAAGRIPT 97
DB 1 MTFSSGLRARPANGTAVQCMGDTAVMTAVKAKPNPQGFMPVVDYRLKNAASGRIPM 60

QY 98 NYLAREVGTSDKELITRSIRIRPLFPAGYFPDTQVLCMLAVDGVNEVDVLAINGAS 157
DB 61 NFMREBELGSEKELISARLIDRSIRPLFKDYRTETQVLCMLAMDVAHSPDLAINAAS 120

QY 158 VALSLSDIPMNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNLVYAGAPKSOIWLKASAE 217
DB 121 MALSLSDIPMNGPVGAVRIGIIDGEVYVNPTRKELQTSQDLVVSATKONLVNMLEKGN 180

QY 218 NILOQDFCHAIKVGKVTQOIIQGIQOLVKETGVTKTPKLTFTSPSEIVYTKHLMER 277
DB 181 VVLQODLLKAIKQGTREAOPIIHEIBRLQKAYGKQKEVEVAADVDELGAVASWCEMR 240

QY 278 LYAVFTYEHDKVSRDEAVNKIRLDTEBQLEKFPPEADPYEIIISFNVAKEVRSYVLN 337
DB 241 LREIFQOSTHDKMSRDNAVNEVRSNVIDKWSFPPDEPSLITQFOFQTSITIRRELIPE 300

QY 338 EYKRCQDRDLTSLNVSCGEVMPFKTLHGSALFORQOTQVLTCTFEDSLSGIKSDVITA 397
DB 301 RGLRCQDRDIDQLNNISQVDMKPLHGSALFORQOTQVLTCTFEDSLSGIKSDVITA 359

QY 398 IN--GIKDKNFMILYEPPEVATNIEIGVTLARRBELGALAEKALYVPIPRDEPPIRV 455
DB 360 LDSSGLKAKNFMILYEPPEVATNIEIGVTLARRBELGALAEKALYVPIPRDEPPIRV 419

QY 456 TSEVLENSGSSMASACGSLALMDSGVPISSAVAGVIGVTK--TDEPEKEIIDYRLLT 514
DB 420 TSEVLENSGSSMASACGSLALMDSGVPISSAVAGVIGVTK--TDEPEKEIIDYRLLT 479

```

```

QY 515 DIIGIEDYNDMDKINAGTKNGITATAGADIKLPGIPIKIYMEALIQASVAKKEILQIMNK 574
DB 480 DIIGIEDYNDMDKINAGTKNGITATAGADIKLPGIPIKIYMEALIQASVAKKEILQIMNK 539

QY 575 TISKPRASRKXNGVYVETVOVPLSKRAKFPVPGGIVNLKLUQAEIGVTISQVDETSVPA 634
DB 540 AIREPRRYKPSWSESTLTYVEPOORALIGPGLHMKRIYLETGSTLTAVDEHFNVFA 599

QY 635 PTPSVMEADPFIETICKDDQEOEFGAVYATITETIRDTGVWVKVLPNMTAVLHNTQ 694
DB 600 PSCAMMEAKBELLEGVYKERVDPLEFGIYITAKITELRDTGVWVILYPSMPALHNSQ 659

QY 695 LDNERL 700
DB 660 LDQSKI 665

RESULT 11
Q8IH29 PRELIMINARY; PRT; 717 AA.
AC Q8IH29
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE GM16802P (Fragment).
GN ORFNames=CG11337;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
NCBI_Locus=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorese V., Dresnek D., Farfan D., Frise E.,
RA Gonzalez M., Guarin H., Krommiller B., Li P., Lao G.,
RA Miranda A., Mungall C.J., Munoz J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Cejnkner S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 KH domain.
CC EMBL; BT001457; AAN71212.1; -.
DR FlyBase; FBgn0039846; CG11337.
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; 3_ExonNase.
DR InterPro; IPR004087; KH.
DR InterPro; IPR004088; KH type 1.
DR InterPro; IPR008994; Nucleic_acid_ob.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH_1; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PSS0084; KH_TYPE_1; 1.
DR PROSITE; PSS0126; S1; 1.
DR NON TER 1
SQ SEQUENCE 717 AA; 79403 MW; 6C1CA7ABC1714D02 CRC64;

Query March 49.5%; Score 1761; DB 2; Length 717;
Best Local Similarity 54.4%; Pred. No. 1.3e-96;
Matches 336; Conservative 119; Mismatches 159; Indels 4; Gaps 3;

QY 86 ROKAAAGRIPTVYLRREVSTSKELITRSIRIRPLFPAGYFPDTQVLCMLAVDGV 145
DB 46 RMTFSSGRIIPMNFMRRELGSEKELISARLIDRSIRPLFKDYRTETQVLCMLAMDVA 105

QY 146 NEPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNLVYAGAP 205
DB 106 HSPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEVYVNPTRKELQTSQDLVVSATK 165

```

QY 206 KSGIWMLEASAEKIILOODCSHAIKVGVKTYQOIIIGIOOLVETSVTKRTOQKLFPSPE 265  
 QY 166 QNVLVHEGKGNVLOODDLKAIKQSTRAGQIHEIERLOKAYORQREVBVAEVDPE 225  
 QY 266 IVKTYHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEBOLKEKPEADPYEIIIESPNV 325  
 Db 226 LGKAVNSMCEMRJREIFODSTHD KMSRDNAVNEVSNNVIDKWSGFPOTEPSSLITTEBPNQ 285  
 QY 326 VAKVEPFSIYLANEYKRCDDRDLTLSLRNVSCEYDMFKTLHGSAIIPORQTOYLCTVTPDSL 385  
 Db 286 TSRTIREBELIFENGACDGDYDQLRNISQYDMKFKPLHGSALIPQRGTOYFCVYSIDSQ 345  
 QY 386 ESGIKSDOYITAIN-GIKDKFMILHYEPPYATNIEKVTYGLNRELIGHALAEKALYP 443  
 Db 346 ESAMKIDS-LAALDSGLKAKNFMLHYEPPYATGEVGRIGVGREREMGHALERSLLP 404  
 QY 444 VIRADPEPTRRVSLEYLSENGSSSMAACGSLALMDGVPISASVNGVALGLVTK-TDP 502  
 Db 405 TLPRDPEFVRLTSBYLVESNGSSSMAVCGSSLAMLDGVPISASVNGVALGLVTKEND 464  
 QY 503 EKGEIEDRYLLTDJLIGIEDYNGDMDFKIGTKNGKITLALQADIKPGLPIKTIWBAIOAS 562  
 Db 465 DTKHLDQOYRILTDJLIGIEDYMGDMKVAGTRKGFALQADIKPGLPIKTIWBSLQAT 524  
 QY 563 VAKKELIQTNNKTIKSPRAKRENGVYETVOYPLSKRAKVPYGPQNYLKLQAFETVTI 622  
 Db 525 DAKSKLIDIMSEAIRPFRKYPKESWPSVSETLVPEPOQROLIGBGLMKRIYLETSTSL 584  
 QY 623 SOVDEETFSVAFATPSPVMEHARDPIELICDQDOOLEFGAVYTTATITIEDIGVMFKLY 682  
 Db 585 TAYDETHFNVAFSQAAMEAKELEIGYVWKRVPYDLDFGGYITAKITIELADTGVMIIV 644  
 QY 683 PNMTVALINTQDNERL 700  
 Db 645 PSMPPALHNHSDORKI 662

RESULT 12			
09S7G6			
ID	09S7G6	PRELIMINARY;	PRT; 991 AA.
AC	09S7G6;		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, last sequence update)		
DT	05-JUL-2004 (TEMBLrel. 27, last annotation update)		
DE	Polynucleotide phosphorylase precursor.		
GN	Name=Epp; Synonyms=TL5N1_70;		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RP	Kim Y.J., Zhou D.X., Mache R.;		
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RA	Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,		
RA	Rudd S., Lemcke K., Mayer K.F.X.;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RM	[3]		
RP	SEQUENCE FROM N.A.		
RP	EU Arabidopsis sequencing project;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; Y14686; CAB43865.1; -.		
DR	EMBL; AL163792; CAB87625.1; -.		
DR	EMBL; Y14685; CAB43864.1; -.		
DR	PIR; T48631; T48631.		
DR	HSSP; P05055; ISRO.		
DR	GO: GO:0000175; F:3'-5'-exoribonuclease activity; IEA.		
DR	GO: GO:0003723; P:RNA binding; IEA.		
DR	GO: GO:0006396; P:RNA processing; IEA.		

DR	InterPro	IPR001247; 3	ExoRNase.
DR	InterPro	IPR004087; KH.	
DR	InterPro	IPR004088; KH_type_1.	
DR	InterPro	IPR008994; NucIc_acid_OB.	
DR	InterPro	IPR003029; SL.	
DR	Pfam	PF00013; KH_1.	
DR	Pfam	PF03726; NPaase; 1.	
DR	Pfam	PF01138; RNase_PH; 2.	
DR	Pfam	PF03725; RNase_PH_C; 2.	
DR	Pfam	PF00575; SL; 1.	
DR	SMART	SM00332; KH; 1.	
DR	SMART	SM00316; SL; 2.	
DR	PROSITE	PS0126; SL; 1.	
KW	transalt peptide.		
FT	TRANSIT	1	Potential.
FT	CHAIN	41	Potential.
FT	CHAIN	991	Potential.
SEQUENCE	991 AA;	107771 MW;	9557E259806D1A5 CRC64;
Query Match		38.4%;	Score 1365.5; DB 2; Length 991;
Best Local Similarity		41.7%;	Pred. No. 8.6e-73;
Matches 287;	Conservative 138;	Mismatches 231;	Indels 33; Gaps 11

QY	24	SAGSRARV----	AVDLGRKKEIESGKILARPADSAVVOQSODTAVMTVAIBKTXKSPSOF	78
Db	44	SAGRIKIESFKEEBEVSGRVVS	PFQKILARFANGSVVLGMDETKVLSVTCAKTXSPDRF	103
QY	79	MPLYVDYRQKAAAAGR	PTNYLREVEGTSDEKELTSRILDSIRPLPAGFYPTDVLN	138
Db	104	LPLTVDOYOEKQAOGL	LPNTYMRREBGAPEKRELLCGRLIDRIPFLPFTGVEHQWAS	163
QY	139	LLAADVNEPQVLA	INGASVALSJDIPANQPVCAVRIGIDGEEVVAWPTREKMSSTLN	198
Db	164	VLSSDGQODPPI	LAMNASSALMLSDVPWGPFIGVIRIGICQGFVNAPTMDLSSSDLN	223
QY	199	LVAAGAPKSQIWM	LEASAEMLDODFCHAKV---GVKYTOQIIOGIIQOLVKETGVTKR	254
Db	224	LIVA-CTRDKMTM	DIVDOSREISEKDLAALTLAPAEAKYULDPQI---RLAEKAGKQCK	278
QY	255	TPQGLFTPSP	EIVKYTHKLAMERYAAPTDEYHOKVSDEAVNKRILRDPTEQLEKKEPDEA	314
Db	279	EYKLSMLSDTK	LEKADLAATRIEASFPTDSYGFKEFGEALMDINIGDVRKVPBEEGQOE	337
QY	315	DPEYIEBFPNAKE	VEFRSIVLNEYKRCGRDLTSLRNVSCEDVMFKTLHGSALFQRCQT	374
Db	338	SLSLPRAVDTR	KVKVRSRMISDGFVRDGHVBEVRICYESHYLPALHGSALFSRQDT	397
QY	375	QVLTCTYFPDS	ESGKSDQVTLANGIKDKRNFMLHYEPFPATNIEIGVTOGLNREIDHG	434
Db	398	QVLTCTYVTLGA	FAEAKQSDSLV---GPEKRFMLHYSPPTCTNVEVGRGGJNRRREVGH	453
QY	435	ALAEKALYVP	PRD--PFFTIRVISEVLESNKSSMASACGSLALMDSGVPISSAGVAV	492
Db	454	TLAEKALLAV	LPPEAPPYTIRINSEVWSDGSTRSMASVCGGSALMDAGIPLARHVAGV	513
QY	493	AIGLVYTKDP	KEGIEDYRLITDLIGEDVNGMDFKIAGTNKGITLALQADIKLPGIPIK	552
Db	514	SVGLITVDV	PESSGRIKYRIYITDLIGEDHIDGMDFKIAGTRDGYTALQLDIKPAGIPLD	573
QY	553	IYMAEIOOAS	VAKKEIIOIMNKTISKPRASHKENGVPVETOVPVLSKRAKFFVGPAGYMLK	612
Db	574	IYVCSLEIA	RAKARQIOLIDHERNINSPRGOGAYSPLRLATIKSNDLSRLTILGPMGVLKR	633
QY	613	KLQAEETL	ISQVDEETFSVPAPTPSVMHAEAR---DFTIEICKDQEOGLEFAGVYTAIT	669
Db	634	KIIEVETGARIS	-IDNGTLTIIVAKNQDVMEKAQEOQVDFII-----GRELVGGVYKGTIV	685
QY	670	TEIRDPDGV	MYKLYIPNMTAVLILHNTQDUNE	698
Db	686	SSIKIEYGA	FVE-PPGQOGLHMLSELHSE	713

RESULT 13  
Q6KA10

```
ID OGA10 PRELIMINARY; PRT; 982 AA.
AC OGA10;
DT 05-JUN-2004 (TEMBLrel. 27, Created)
DT 05-JUN-2004 (TEMBLrel. 27, Last sequence update)
DE Putative polyribonucleotide nucleotidyltransferase.
GN Name-OJ1014_H03.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003980; BAD21450.1; -
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001247; 3 ExoRNase.
DR InterPro; IPR008994; Nucleic_acid_OR.
DR InterPro; IPR003029; S1.
DR Pfam; PF03726; RNase; 1.
DR Pfam; PF01138; RNase PH; 2.
DR Pfam; PF03725; RNase PH_C; 2.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 2.
DR PROSITE; PS50126; S1; 1.
DR Transferrase.
SQ SEQUENCE 982 AA; 107027 MW; 0B05E5A0B684351 CRC64;

Query Match 36.2%; Score 1286.5; DB 2; Length 982;
Best Local Similarity 39.3%; Pred. No. 4.5e-68;
Matches 286; Conservative 138; Mismatches 253; Indels 51; Gaps 13;

QY 8 RRDRALTOLOV---RALMSAGSRAVA-----VDLGNKLEISS 43
DB 17 RRARFPAPLSPVGRAPFLSGAAAEVAQADAPPPPGKATLESFREFEIGRIVISFET 76
QY 44 GKLAEPDGSAAVOSGDTAVMT-AVSKTSPSQFMPVVDYRQKAAAAGRIPTNYLR 102
DB 77 GKMAFPANGSVVISMDDTHVLSVAAKSSRPVDFPLTVDYQEKQYAGVIAPTTYMR 136
QY 103 EVGTSDEKILTSRIIDRSIRPLFPAGYFYDTQV-----LCNLAVDGVNEPVLAINGA 156
DB 137 EGAPKERELLCGRIIDRPIPLPPPGFYHEVQVNNATIIWNVVSSGKODPPVMAANAS 196
QY 157 SVALSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNVVAGAPKSQIVMLEASA 216
DB 197 SAALMSLDIPWNGPIGVIRGRIDGNFVLANPTVDELGSLDLNVYA-CSRDKTLMDVQA 255
QY 217 ENLIQDFCHAIKVGKVTQOIIQGIQOLVKETGVTKTPQKLFPPSPBEIVKTYHKLAME 276
DB 256 REITERDLQGMKLAHAEAVKCINPOLRLAKRAG-KKKKEKYKISLSDSKYKIRTLSEA 314
QY 277 RLVAFTDYHDYKVSDEAVNKKRLDTEBOLKKEKFPADPYEIIIESFNVAKEVFRSIVL 336
DB 315 PIEEVFTDSTYGFKEFGEALENITQSVAKALEECDEDSLKFHKADVTRAKQVIRKRII 374
QY 337 NEYKRCGRDLTSLRNVSCVDMPKTLHGSALFORGOTVLCVTPDLSLEGISD-QVI 395
DB 375 EKGIRVGRQLDEVRPLVCSSSTYPIHGSALFSGDQVQVCTYTL-----GAGDQRL 429
QY 396 TAINGIKDKKFMFLHYEPPPYATNEIGKVTGLNRRELGHGALAEKALYVPVIR-RDPPTI 453
DB 430 DSIIGPPTKRFMLHYSPFPFISINEVAKRGGLNRREVGHGTLAEKALAVLPPSEGEFPPTV 489
QY 454 RYTSVEVLESNSSSSMASCGSLALMDSGVPISSAVGVALGVTNKDPEKGELEDRL 513
DB 490 RVNSEVWASDSTSMASVCGGSMALMDAGIVRSHVAVSVGLVSEVDQTTGDISSTRIL 549
QY 514 TDIIGIEDYNDMPFKLAGTKGITALQADIKLPGIPIKIWEAIIQASVAKKEILQIMN 573
DB 550 TDIIGLEDHLDGMDPKLAGTRGITATQDLIKPAGIPLDITCESLERPAKRNQILDRMD 609
```

```
QY 574 KTISKPRASRKNENGPIVEITQVPLSKRAKFGVPGYNLKKQQAETGVITISQVDEFTFSVF 633
DB 610 QEISSARAFVNDSSPRLATISFSSDSLRLKL-----PFRKKIEGETGARVS-VSDGVTIV 664
QY 634 APPPSVWHEARDFITTEICKDDEQQLSEFGAVYATITTEIRDTGVWYKILPMTAVLLHNT 653
DB 665 AKTPQIMDKAKIEKVEFLV-----GREIEVGRYKGVSSIKEXGAFVEFNGGQG-LIHIS 719
QY 694 QLNERLUN 701
DB 720 ELSHDKVS 727

RESULT 14
ID BAD21450 PRELIMINARY; PRT; 982 AA.
AC BAD21450;
DT 01-JUN-2004 (TEMBLrel. 27, Created)
DT 01-JUN-2004 (TEMBLrel. 27, Last sequence update)
DE Putative polyribonucleotide nucleotidyltransferase.
GN OJ1014_H03.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza; Oryza sativa.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone OJ1014_H03.14";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003980; BAD21450.1; -
DR Transferrase.
SQ SEQUENCE 982 AA; 107027 MW; 0B05E5A0B684351 CRC64;

Query Match 36.2%; Score 1286.5; DB 2; Length 982;
Best Local Similarity 39.3%; Pred. No. 4.5e-68;
Matches 286; Conservative 138; Mismatches 253; Indels 51; Gaps 13;

QY 8 RRDRALTOLOV---RALMSAGSRAVA-----VDLGNKLEISS 43
DB 17 RRARFPAPLSPVGRAPFLSGAAAEVAQADAPPPPGKATLESFREFEIGRIVISFET 76
QY 44 GKLAEPDGSAAVOSGDTAVMT-AVSKTSPSQFMPVVDYRQKAAAAGRIPTNYLR 102
DB 77 GKMAFPANGSVVISMDDTHVLSVAAKSSRPVDFPLTVDYQEKQYAGVIAPTTYMR 136
QY 103 EVGTSDEKILTSRIIDRSIRPLFPAGYFYDTQV-----LCNLAVDGVNEPVLAINGA 156
DB 137 EGAPKERELLCGRIIDRPIPLPPPGFYHEVQVNNATIIWNVVSSGKODPPVMAANAS 196
QY 157 SVALSLSDIPWNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNVVAGAPKSQIVMLEASA 216
DB 197 SAALMSLDIPWNGPIGVIRGRIDGNFVLANPTVDELGSLDLNVYA-CSRDKTLMDVQA 255
QY 217 ENLIQDFCHAIKVGKVTQOIIQGIQOLVKETGVTKTPQKLFPPSPBEIVKTYHKLAME 276
DB 256 REITERDLQGMKLAHAEAVKCINPOLRLAKRAG-KKKKEKYKISLSDSKYKIRTLSEA 314
QY 277 RLVAFTDYHDYKVSDEAVNKKRLDTEBOLKKEKFPADPYEIIIESFNVAKEVFRSIVL 336
DB 315 PIEEVFTDSTYGFKEFGEALENITQSVAKALEECDEDSLKFHKADVTRAKQVIRKRII 374
QY 337 NEYKRCGRDLTSLRNVSCVDMPKTLHGSALFORGOTVLCVTPDLSLEGISD-QVI 395
DB 375 EKGIRVGRQLDEVRPLVCSSSTYPIHGSALFSGDQVQVCTYTL-----GAGDQRL 429
QY 396 TAINGIKDKKFMFLHYEPPPYATNEIGKVTGLNRRELGHGALAEKALYVPVIR-RDPPTI 453
```

Db 430 DSIYGPTEKFMFLYSPPEPSINEVAKRGGINREVEGHGTLAEKALLAVLPBEGEPFTV 489  
QY 454 RYTSVEVSNSSSSMASACGSLALMDSGVPISSAVAGVAILGYTXTPDPEKEIEDYRL 513  
Db 490 RVNSEVVASDOSTIMASVCGSMLMDAGIPVRHVAVGVSGLVSEVDQTTGDISSYRL 549  
QY 514 TDLIGIEDYNGDMDFKAGTNKGITLQADIKLPKIPKIVMEAIQQAASVAKKEILQIMN 573  
Db 550 TDLIGLBDHIGDMDFKAGTRRGITAIQIDIKPAGIPLDITCEGLBARKARNOILDRMD 609  
QY 574 KITSKPAASKRENGPVVETQVPLSKRAKFPVPGVNLKQLQATGVTISQVDETSVF 633  
Db 610 QEISSAARFNDGSSPRLATISFSSDSLRLKLL---FHRKKIEQETGARVS-VSDGVTIV 664  
QY 634 APSPVHAEADFTIEICKDQEOLEFGAVYATITIEIRDTGVWTKLYPMNTAVLHMT 693  
Db 665 AKTPIPDKALKEKVEFLV---GREIEVGRYKGVSVSIKEYGFVFEFNGGQG-LHHS 719  
QY 694 QLDNERLN 701  
Db 720 ELSDKVS 727  
RESULT 15  
Q96SN3 PRELIMINARY; PRT; 745 AA.  
ID Q96SN3;  
AC Q96SN3;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Hypothetical protein BE0003N10.1.  
GN ORFNames=BE0003N10.1;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OC NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Watson R.;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Wilson R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC092690; AAK73855.2; -.  
DR HSSP; Q53597; IASP.  
DR WormPep; BE0003N10.1; CE30235.  
DR GO; GO:0000175; F:3'-5'-exoribonuclease activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0006356; P:RNA processing; IEA.  
DR InterPro; IPR001247; 3 ExonName.  
DR InterPro; IPR004088; KH\_type\_1.  
DR Pfam; PF00013; KH\_1; 1. type\_1.  
DR Pfam; PF03725; nRase; 1.  
DR Pfam; PF01138; RNase\_PH\_2.  
DR Pfam; PF03725; RNase\_PH\_C; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 745 AA; 82881 MW; 45D03070C72ACB16 CRC64;

Query Match 35.9%; Score 1276; DB 2; Length 745;  
Best Local Similarity 42.2%; Pred. No. 1,3e-67;  
Matches 278; Conservative 109; Mismatches 237; Indels 34; Gaps 10;

QY 28 RAAVAVDLGNRKLISGKLAFADGSAVVGSDTAVAVTAVS-KTKPSPSQFMPLVVDY 85  
Db 26 QSAVAVDPGKSGIEKAKHLARFASSTVYASFGDNAVAVATVQKSKGDSGVPLOVEY 85  
QY 86 RQKAAAGRIPTNYLRREVSTSDKEILTSRIIDRSIRPLPAGYFYDTQVLGNLLAVDGV 145

Db 86 RPSASAIGRIPTEFLRELSQSDNEILISPAIDRSIRPLP-ONSVEYQITCKPLADEN 144  
QY 146 NEPDVLAINGASVALSIDIPMNGPYGAVRIG-IIGEVYVNFTRKMSSTLNLVAGA 204  
Db 145 ADQIMGICINASTALQISSAAYNGPLALRVATAGDFHVNVNTOBELRASINLIVAMR 204  
QY 205 PKSQIVMLEASAENILQODECHAIKVGVKXQQIIOGIIQQLVKEVTKTPQKLFPPSP 264  
Db 205 KHEKTVMIEDGRESSAEHLEHALDVAFRHVAKLHEAMEQLTAE-----PKD 251  
QY 265 EIV-----KTHKLAMRLYAVFTDYEHDSVDEAVNKRRLDTEQLKEX-PEAD 315  
Db 252 ELASEDFSGLERLLEETARERIVYITDAGHDKISDMETKAL---PEELCAKAFOTCE 308  
QY 316 PYEIESFNVVAKEVPSIYLANEYKRCDSGLTSLNVSCEVDMFKTLHGSALFORGOTO 375  
Db 309 KDAIYRYSTLVKKVLRDTTLRTGIRKCDGRPRPEFRPITIHVMYKTLHCCSIPOGQIQ 368  
QY 376 VLCTVTEDSLSEGIKSDQVITAINGIKDNFMFLHYEPPPYATNIEIGKVTGLNRELGHA 435  
Db 369 VMSTVTFDSPAALFHPDSVAQLIGSQKKSFMFLHYEPPGPAINEFRTBSLNREIGHA 428  
QY 436 LAEKALYPVIRPDPPTIRTVTSVLESNGSSMASACGSLALMDSGVPISSAVAGAIG 495  
Db 429 LAEKSILNLPADFPYATRLACQVLESNGSSMASVCGSLALPDAGVPMKMAAGVAIG 488  
QY 496 LVT-KTDPKEGEIEDYRLTDLIGIEDYNGDMDFKAGTNKGITLQADIKLPKIPKIY 554  
Db 489 LIDDEAPE-----TKRVLVDIIGIEDYAGDMDFKAGTMDGFTAAQLDVKNGLTRQL 544  
QY 555 MEAIQQAASVAKKEILQIMNTTISKPPASRENGPVVETQVPLSKRAKFPVPGVNLKQL 614  
Db 545 TESIQARABAGIDHVLQMSVWRDRPREQFKPTPIIIOQMKRIEPRKRTTLFRNNGYCKLI 604  
QY 615 QAGTGVTIQVDETSVFAPTPSVHAEADFTIEICKDQEOLEFGAVYATITIEI 672  
Db 605 EAETGVKISADEAHISLQDKEKLOKAMDMDVLSNSTIDFAGSIVQAEIYEV 662

Search completed: January 28, 2005, 19:34:54  
Job time : 208 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 28, 2005, 19:16:56 ; Search time 165 Seconds  
(without alignments)  
1532.752 Million cell updates/sec

Title: US-09-907-907a-42  
Perfect score: 3557  
Sequence: 1 DGFPLPRRDRALTLQVRA.....TAVLLHNTQLDNERLITLLP 705

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues  
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3557	100.0	705	3 AAB08732	AAB08732 Amino aci
2	3556	100.0	722	7 ADC39094	ADC39094 Novel hum
3	3498	98.3	783	5 ABP68432	ABP68432 Human pol
4	3030	85.2	675	5 ABG30875	ABG30875 Human pol
5	2289	64.4	540	7 ADC39092	ADC39092 Novel hum
6	2285	64.2	504	4 AAB92684	AAB92684 Human pro
7	2285	64.2	504	7 ADJ69985	ADJ69985 Human hea
8	1982	55.7	899	4 ABG17275	ABG17275 Novel hum
9	1982	55.7	899	4 ABG08546	ABG08546 Novel hum
10	1982	55.7	899	7 ADE08997	ADE08997 Novel pro
11	1964	55.2	476	7 ADE07933	ADE07933 Novel pro
12	1917	53.9	439	4 ABG17276	ABG17276 Novel hum
13	1842.5	51.8	748	4 ABB58546	ABB58546 Drosophi
14	1228.5	34.5	541	4 ABG08547	ABG08547 Novel hum
15	1204.5	33.9	703	6 ABU24041	ABU24041 Protein e
16	1181.5	33.5	696	4 AAU33947	AAU33947 Staphyloc
17	1181.5	33.5	698	4 AAU36732	AAU36732 Staphyloc
18	1190.5	33.5	698	6 ABU16489	ABU16489 Protein e
19	1190.5	33.5	698	6 ABM73120	ABM73120 Staphyloc
20	1190.5	33.5	698	7 ADD26218	ADD26218 Staphyloc
21	1190.5	33.5	698	7 ADD26220	ADD26220 Staphyloc
22	1165.5	32.8	701	6 ABU43024	ABU43024 Protein e
23	1165.5	32.7	706	5 ABP40063	ABP40063 Staphyloc
24	1161.5	32.6	711	6 ABU43664	ABU43664 Protein e
25	1161	32.6	711	6 ABU47605	ABU47605 Protein e

26	1158	32.6	721	4 AAU38175	AAU38175 Salmonell
27	1157	32.5	729	6 ABU33310	ABU33310 Protein e
28	1156	32.5	1034	2 AAU03792	AAU03792 S. aureus
29	1155	32.5	703	6 ABU25139	ABU25139 Protein e
30	1155	32.5	734	4 AAU34719	AAU34719 E. coli c
31	1155	32.5	734	6 ABU28778	ABU28778 Protein e
32	1150.5	32.3	719	7 ABO67122	ABO67122 Klebsiell
33	1147	32.2	711	6 ABU31568	ABU31568 Protein e
34	1146.5	32.2	713	6 ABU19845	ABU19845 Protein e
35	1146	32.2	705	6 ABU50011	ABU50011 Protein e
36	1140	32.0	709	6 ABU40849	ABU40849 Protein e
37	1140	32.0	720	7 ADP05453	ADP05453 Bacteri
38	1135.5	31.9	719	6 ABU23367	ABU23367 Protein e
39	1135	31.9	712	6 ABU28271	ABU28271 Protein e
40	1134	31.9	709	4 AAU35396	AAU35396 Haemophil
41	1134	31.9	709	6 ABU30185	ABU30185 Protein e
42	1131.5	31.8	771	6 ABU20913	ABU20913 Protein e
43	1129.5	31.8	708	7 ADH88109	ADH88109 Enterococ
44	1128.5	31.7	704	4 AAU35289	AAU35289 Enterococ
45	1128.5	31.7	704	6 ABU14497	ABU14497 Protein e

## ALIGNMENTS

RESULT 1  
AAB08732 standard; protein, 705 AA.  
ID AAB08732  
XX  
AC AAB08732;  
XX  
DT 02-JAN-2001 (first entry)  
XX  
DB Amino acid sequence of a human OLD-35 polypeptide.  
XX  
XX OLD-35; OLD-64; OLD-137; OLD-139; OLD-142; OLD-175; cancerous phenotype;  
XX cellular senescence; terminal differentiation; growth suppression;  
XX aging process; type I interferon; cancer cell; tissue regeneration; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200046391-A2.  
XX  
PD 10-AUG-2000.  
XX  
PP 02-FEB-2000; 2000MO-US002920.  
XX  
PR 02-FEB-1999; 99US-00243277.  
XX  
PA (UYCO ) UNIV COLUMBIA NEW YORK.  
XX  
XX Fleher PB, Leszczynska M;  
XX  
XX WPI; 2000-532905/48.  
XX N-PSDB; AAA64608.  
XX  
PT Novel isolated nucleic acid encoding an OLD-35 or OLD-64 protein useful  
PT in the treatment and detection of e.g. cancer and diseases involving  
PT cellular senescence.  
XX  
XX  
PS Disclosure; Fig 9B; 115pp; English.  
XX  
XX The specification describes OLD-35, OLD-64, OLD-137, OLD-139, OLD-142 and  
XX OLD-175 proteins. The OLD nucleic acids are useful for reversing the  
XX cancerous phenotype of a cancer cell, determining if a cell is senescent,  
XX growth arrested or terminally differentiated. They are also useful for  
XX reversing the aging process in a cell and degrading specific RNAs in a  
XX cell. The genes may also be used as a diagnostic indicator of cellular  
XX senescence, terminal differentiation and/or growth suppression and as a  
XX marker to identify drugs or small molecules that will induce or inhibit  
XX cellular senescence or terminal differentiation and type I interferons.  
XX The combination of Old-35 with other interacting proteins is useful for  
XX targeting the differentiation of specific cells. Old-35 can be used to

CC selectively stabilize specific mRNAs containing adenoviral rich 3' UTRs.  
 CC The Old proteins are useful for reversing the cancerous phenotype of a  
 CC cancer cell and inhibiting the growth of a cancer cell. They are also  
 CC useful for regenerating tissue. The present sequence represents an Old-35  
 CC polypeptide

XX Sequence 705 AA:

Query Match 100.0%; Score 3557; DB 3; Length 705;

Best Local Similarity 100.0%; Pred. No. 1,7e-301; Mismatches 0; Indels 0; Gaps 0;

Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPELPRDRALQLOVRAIMSSAGSRVAVDJGNRKLEISGSKLARFADGSAVVOGSD 60  
 DB 1 DGPELPRDRALQLOVRAIMSSAGSRVAVDJGNRKLEISGSKLARFADGSAVVOGSD 60  
 QY 61 TAVWTVSKTKPSQPMPLVDYRQAAAGRIPTNYLRKRGTSDEKELTSTRIDRS 120  
 DB 61 TAVWTVSKTKPSQPMPLVDYRQAAAGRIPTNYLRKRGTSDEKELTSTRIDRS 120  
 QY 121 IRPLPAGYFYDQVLCNLAVDGVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180  
 DB 121 IRPLPAGYFYDQVLCNLAVDGVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180  
 QY 181 GEYVNPTRKEMSSSTLNVVAGAPKSOIWMLESAENILQDPFGHAIKGVKTTQOIIO 240  
 DB 181 GEYVNPTRKEMSSSTLNVVAGAPKSOIWMLESAENILQDPFGHAIKGVKTTQOIIO 240  
 QY 241 GIQOLVETGVTKTPOKLTFTSPSEIVKYTHKLMEERYANFTDYEHDKVRDAVVKIR 300  
 DB 241 GIQOLVETGVTKTPOKLTFTSPSEIVKYTHKLMEERYANFTDYEHDKVRDAVVKIR 300  
 QY 301 LDTEEOQKEKEPEADPEYELIESFNVAKEVRSIVLNEYKRCDSRDLTSLANVCEVDMF 360  
 DB 301 LDTEEOQKEKEPEADPEYELIESFNVAKEVRSIVLNEYKRCDSRDLTSLANVCEVDMF 360  
 QY 361 KTLHGSALEFORGQOVLCCTVTPDSLESGIKSDQVITAINGIKDXNFMILHYEPYATNEI 420  
 DB 361 KTLHGSALEFORGQOVLCCTVTPDSLESGIKSDQVITAINGIKDXNFMILHYEPYATNEI 420  
 QY 421 GKTVGLNRRELGHGALAEKALYPIPRDPERTIRVTSFVLESNSSSMAACGGSLALMD 480  
 DB 421 GKTVGLNRRELGHGALAEKALYPIPRDPERTIRVTSFVLESNSSSMAACGGSLALMD 480  
 QY 481 SGVPISSAVAGVALGVTCTDPEKGEIEDYRLLDIIGEDYNDMDPFKLAGTKGITAL 540  
 DB 481 SGVPISSAVAGVALGVTCTDPEKGEIEDYRLLDIIGEDYNDMDPFKLAGTKGITAL 540  
 QY 541 QADIKLPGIPIKIYWEAIOQASVAKKEILOIMNTTISKPRASRENGPVETVOVPLSKR 600  
 DB 541 QADIKLPGIPIKIYWEAIOQASVAKKEILOIMNTTISKPRASRENGPVETVOVPLSKR 600  
 QY 601 AKFVPGSGYNKKIOAETGVVISOVDETSVFAPTPSVHMEARDFTIEICKDOEOOLE 660  
 DB 601 AKFVPGSGYNKKIOAETGVVISOVDETSVFAPTPSVHMEARDFTIEICKDOEOOLE 660  
 QY 661 FGAVYTAITITEIRDTGVWVKLYPMNTAVLANTOLDNERLNIILP 705  
 DB 661 FGAVYTAITITEIRDTGVWVKLYPMNTAVLANTOLDNERLNIILP 705  
 RESULT 2  
 ADC39094  
 ID ADC39094 standard; protein; 722 AA.  
 AC ADC39094;  
 DT 18-DEC-2003 (first entry)  
 DE Novel human NOVX polypeptide SEQ ID NO: 36.  
 KW antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;  
 KW nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;

KW antiaesthetic; antiinflammatory; hypotensive; antiarteriosclerotic;  
 KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;  
 KW lymphoma; uerous cancer; prostate cancer; dyslipidemia; anorexia;  
 KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;  
 KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;  
 KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;  
 KW Albright hereditary osteodystrophy.

XX Homo sapiens.

OS WO2003010327-A2.

PN 06-FEB-2003.

PF 02-MAY-2002; 2002WO-US014199.

XX 02-MAY-2001; 2001US-0288063P.  
 PR 03-MAY-2001; 2001US-0288395P.  
 PR 07-MAY-2001; 2001US-0289087P.  
 PR 09-MAY-2001; 2001US-0289817P.  
 PR 09-MAY-2001; 2001US-0289818P.  
 PR 11-MAY-2001; 2001US-0290194P.  
 PR 14-MAY-2001; 2001US-0290753P.  
 PR 15-MAY-2001; 2001US-0291181P.  
 PR 16-MAY-2001; 2001US-0291243P.  
 PR 18-MAY-2001; 2001US-0292001P.  
 PR 21-MAY-2001; 2001US-0292374P.  
 PR 22-MAY-2001; 2001US-0292587P.  
 PR 23-MAY-2001; 2001US-0293107P.  
 PR 25-MAY-2001; 2001US-0293747P.  
 PR 29-MAY-2001; 2001US-0294109P.  
 PR 29-MAY-2001; 2001US-0294110P.  
 PR 30-MAY-2001; 2001US-0294434P.  
 PR 31-MAY-2001; 2001US-0294827P.  
 PR 12-JUL-2001; 2001US-0304879P.  
 PR 31-JUL-2001; 2001US-0308901P.  
 PR 14-AUG-2001; 2001US-0312270P.  
 PR 17-AUG-2001; 2001US-0313416P.  
 PR 10-SEP-2001; 2001US-0318463P.  
 PR 27-SEP-2001; 2001US-0325683P.  
 PR 18-OCT-2001; 2001US-0330292P.  
 PR 28-NOV-2001; 2001US-0333873P.  
 PR 03-DEC-2001; 2001US-0336909P.  
 PR 03-DEC-2001; 2001US-0337552P.  
 PR 21-FEB-2002; 2002US-0359245P.  
 PR 01-MAY-2002; 2002US-00136826.  
 XX (CUBA-) CURAGEN CORP.  
 PA Miller CE, Kekuda R, Maljankar UM, Li L, Pena CEA, Spytek KA;  
 PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zethusen BD;  
 PI Pottarajan M, Anderson DW, Mezes PS, Peyman JA, MacDougall JR;  
 PI Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;  
 PI Edinger SR, Ellerman K;  
 XX WPI; 2003-239445/23.  
 DR N-PSDB; ADC39093.  
 XX New NOVX polypeptides and polynucleotides, useful in gene therapy,  
 PT particularly for treating or preventing a syndrome associated with a  
 PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,  
 PT hypertension or hemophilia.  
 PS Claim 1; SEQ ID NO 36; 748bp; English.  
 XX The invention relates to new isolated NOVX polypeptides, the genes  
 CC encoding them or sequences having at least 95% identity to the amino acid  
 CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,  
 CC particularly in the manufacture of a medicament for treating a syndrome  
 CC associated with a human disease, which includes a pathology associated  
 CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for  
 CC treating, preventing or alleviating pathology associated with NOVX  
 CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and

CC polypeptide are especially useful for treating or preventing e.g. diabetes, obesity, cancers (e.g. lymphoma, uterine cancer or prostate cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease, Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia, graft-versus-host disease or Albritght hereditary osteodystrophy. The DNA encoding the protein is useful in gene therapy for treating the above conditions. These are also useful in developing powerful assay system for functional analysis of various human disorders, as well as in diagnostic applications. This sequence represents one of the NOVX proteins of the invention.

CC  
XX  
SQ Sequence 722 AA;

Query Match 100.0%; Score 3556; DB 7; Length 722;  
Best Local Similarity 99.9%; Pred. No. 2.1e-301;  
Matches 704; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGPELLPRDRALVQLQVRALMSAGRAVAVDLGNRLKLEISSGKLARFADGSAVVOGSD 60  
DB 18 DGPELLPRDRALVQLQVRALMSAGRAVAVDLGNRLKLEISSGKLARFADGSAVVOGSD 77  
QY 61 TAVWTVAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVGTSDEKILTSRIIDS 120  
DB 78 TAVWTVAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVGTSDEKILTSRIIDS 137  
QY 121 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNNGPGAVRIGIID 180  
DB 138 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNNGPGAVRIGIID 197  
QY 181 GEYVNPTRKEMSSSTLNLVAGAPKSIQVWLEASAEINLQODFCHAIKVGKYYTQIIO 240  
DB 198 GEYVNPTRKEMSSSTLNLVAGAPKSIQVWLEASAEINLQODFCHAIKVGKYYTQIIO 257  
QY 241 GIQQLVETGYTKTPQKLTFTSPSEIVKYTHKLANERLYAVFTDYEHDKVSRDEAVNKIR 300  
DB 258 GIQQLVETGYTKTPQKLTFTSPSEIVKYTHKLANERLYAVFTDYEHDKVSRDEAVNKIR 317  
QY 301 LDTEOLKEKEPEADPYEIIISFNVAKEVRSIYLANEYKCDGDLTSLANVCEVDMF 360  
DB 318 LDTEOLKEKEPEADPYEIIISFNVAKEVRSIYLANEYKCDGDLTSLANVCEVDMF 377  
QY 361 KTLHGSAALFORGQTOVLCVTFPDSLSEGIKSDQVITAINGIKDKFMHYEPFYATNEI 420  
DB 378 KTLHGSAALFORGQTOVLCVTFPDSLSEGIKSDQVITAINGIKDKFMHYEPFYATNEI 437  
QY 421 GKVTGLNRRELGHGALAEKALYPVLPDPTIRVTSEVLESNGSSMASACGSLALMD 480  
DB 438 GKVTGLNRRELGHGALAEKALYPVLPDPTIRVTSEVLESNGSSMASACGSLALMD 497  
QY 481 SGVRISSVAVAGVAGLVTKTPPEKGEIDRYLLDILGIEDYNGMDKIKAGTKGITAL 540  
DB 498 SGVRISSVAVAGVAGLVTKTPPEKGEIDRYLLDILGIEDYNGMDKIKAGTKGITAL 557  
QY 541 QADIKLPEIPKIKIWEAIIQASVAKKEIIQIMNKTIISKPRASRKENGVEVTVQVPLSKR 600  
DB 558 QADIKLPEIPKIKIWEAIIQASVAKKEIIQIMNKTIISKPRASRKENGVEVTVQVPLSKR 617  
QY 601 AKFVPGGVNKLKCLQAEITGVITISQVDETFVSFAFPTSPVMEHARDPITEICKDOEQGLE 660  
DB 618 AKFVPGGVNKLKCLQAEITGVITISQVDETFVSFAFPTSPVMEHARDPITEICKDOEQGLE 677  
QY 661 FGAVYTTATITIRDTGVWVKIYPNNTAVLANTQIDNRLNILLP 705  
DB 678 FGAVYTTATITIRDTGVWVKIYPNNTAVLANTQIDNRLNILLP 722

RESULT 3  
ID ABP69432 standard; protein; 783 AA.  
XX AC ABP69432;  
XX

DT 20-JAN-2003 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 1479.  
XX

Human, genome mapping; gene therapy; food supplement; virus; fungus; cell-proliferative disorder; neurodegenerative disease; bacterial; Parkinson's disease; Alzheimer's disease; autoimmune disease; multiple sclerosis; diabetes; genetic disorder; wound; burn; infection; arthritis; cytotoxic; immunomodulator; neotropic; neuroprotective; antiparkinsonian; antidiabetic; immunosuppressive; dermatological; haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide; antichratic.

XX Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

PI Wehrman T, Wang J, Wang D, Drmanac RT;

DR WPI; 2002-759812/82.

DR N-PSDB; AB211649.

XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet

PT or coagulation disorders.

PS Claim 9; SEQ ID NO 1479; 1012bp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (AB211119-  
CC AB212066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP68849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 783 AA;

Query Match 98.3%; Score 3498; DB 5; Length 783;  
Best Local Similarity 98.9%; Pred. No. 2.9e-296;  
Matches 692; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 DGPELLPRDRALVQLQVRALMSAGRAVAVDLGNRLKLEISSGKLARFADGSAVVOGSD 60  
DB 18 DGPELLPRDRALVQLQVRALMSAGRAVAVDLGNRLKLEISSGKLARFADGSAVVOGSD 77  
QY 61 TAVWTVAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVGTSDEKILTSRIIDS 120  
DB 78 TAVWTVAVSKTKPSQSPQFMPVLVDYRQKAAAGRIPTNYLRREVGTSDEKILTSRIIDS 137  
QY 121 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNNGPGAVRIGIID 180  
DB 138 IRPLFPAGYFYDTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPNNGPGAVRIGIID 197



QY	181	GEYVNPFRKEMSSSTLNIVVAGAKPSQIVMLSEASENILQDFCAIKVGVYQYTOIIO	240
Db	198	GEYVNPFRKEMSSSTLNIVVAGAPKSQIVMLSEASENILQDFCAIKVGVYQYTOIIO	257
QY	241	GIQQLVKEGTGYTKRFPQKLFTPSPPEIVKTKHKLAMERLYAVFTDYEDHKVSRDEAVNKR	300
Db	258	CIQQLVKEGTGYTKRFPQKLFTPSPPEIVKTKHKLAMERLYAVFTDYEDHKVSRDEAVNKR	317
QY	301	LDTEBQLKEKPEADPYEIIIESFNVAVAKEVFRSIVLNEYKRCDDGLTSLRNVCEVDMF	360
Db	318	LDTEBQLKEKPEADPYEIIIESFNVAVAKEVFRSIVLNEYKRCDDGLTSLRNVCEVDMF	377
QY	361	KLTHGSALFQRQQTQVLCVPTFDSLSESGIKSPQVITAINGIDKNNMLHYEPPEVATNEI	420
Db	378	KLTHGSALFQRQQTQVLCVPTFDSLSESGIKSPQVITAINGIDKNNMLHYEPPEVATNEI	437
QY	421	GKVTGLNRELGHGALAEKALPVIPRDPFTIRVTSVLESGSSSMASACGSLALMD	480
Db	438	GKVTGLNRELGHGALAEKALPVIPRDPFTIRVTSVLESGSSSMASACGSLALMD	497
QY	481	SGVPISSAVAGVAGLVTKTDPDEKGEIEDYRLITDILGIEDYNGDMFKIAGTNKGITLAL	540
Db	498	SGVPISSAVAGVAGLVTKTDPDEKGEIEDYRLITDILGIEDYNGDMFKIAGTNKGITLAL	557
QY	541	QADILPGIPRIKIWEATIQOASVAKKEILQIMNKTSKRPRASRKENGPVVETQVPLSKR	600
Db	558	QADILPGIPRIKIWEATIQOASVAKKEILQIMNKTSKRPRASRKENGPVVETQVPLSKR	617
QY	601	AKFVPGGVGNLKKLQAEFTGVTIISOVDEETFSVFAPTPSVMHEARDEITTEICKDOEQOLE	660
Db	618	AKFVPGGVGNLKKLQAEFTGVTIISOVDEETFSVFAPTPSVMHEARDEITTEICKDOEQOLE	677
QY	661	FGAVVTATITTEIRDTGVWVKLYPNNTAVULIANTOLDNEKL 700	
Db	678	FGAVVTATITTEIRDTGVWVKLYPNNTAVULIANTOLDNEKL 717	
RESULT 4			
ABG30875	ID	ABG30875 standard; protein; 675 AA.	
XX	AC		
XX	ABG30875;		
XX	DT	21-OCT-2002 (first entry)	
DE	XX	Human polynucleotide phosphorylase 74.25.	
DE	XX	Human; enzyme; polynucleotide phosphorylase 74.25; malignant tumour;	
KW	XX	haemopathy; human immunodeficiency virus infection; HIV;	
KW	XX	immunological disease; inflammation.	
OS	XX	Homo sapiens.	
XX	NN	CN1341720-A.	
XX	PD	27-MAR-2002.	
XX	PP	05-SEP-2000; 2000CN-00119892.	
XX	PR	05-SEP-2000; 2000CN-00119892.	
XX	PA	(SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.	
XX	PI	Mao Y, Xie Y;	
XX	DR	WPI; 2002-501203/54.	
XX	DR	N-PSDB; ABR89228.	
PT	XX	New polypeptide-polynucleotide phosphorylase 74.25 for treating malignant	
PT	XX	tumour, hemopathy, human immunodeficiency virus infection, immunological	
XX	XX	disease and various inflammations.	

PS	Claim 1; page 25-26 (Disclosure), 33pp; Chinese.
XX	
CC	The present invention discloses a new polypeptide-polynucleotide
CC	phosphorylase 74.25, a polynucleotide encoding the polypeptide and a
CC	method for producing the polypeptide using DNA recombination technology.
CC	The invention also discloses a method for curing several diseases, such
CC	as malignant tumour, haemopathy, human immunodeficiency virus (HIV)
CC	infection, immunological disease and various inflammations by using the
CC	polypeptide. The invention also discloses an antagonist for resisting the
CC	polypeptide and its therapeutic action, and also discloses the
CC	application of the polynucleotide encoding the new polynucleotide
CC	phosphorylase 74.25. The present sequence represents the polynucleotide
CC	phosphorylase 74.25
XX	
SQ	Sequence 675 AA;
	Query Match            85.2%; Score 3030; DB 5; Length 675;
	Best Local Similarity   94.5%; Pred. No. 2e-255;
	Matches   602; Conservative   3; Mismatches   4; Indels   28; Gaps   1
QY	64    MTVAVSKTKPSPSQMPPLVVDYRQKAAAGRIPNTYLRRREVTSQDEILTSRIIDRSIRP   123
DB	1    MTVAVSKTKPSPSQGMPPLVVDYRQAAAAAGRLPTNYLRRREIGTSDKEILTSRIIDRSIRP   60
QY	124   LPPAGYPFDTOVLCHNLAVDGVNEBDVLAINGASVALSLSDIPMNGPGPAVARIGIIDGRY   183
DB	61   LPFAGYPFDTOVLCHNLAVDGVNEBDVLAINGASVALSLSDIPMNGPGPAVARIGIIDGRY   120
QY	184   VVNPRFKMSSTLNLVAGA PKSGI VWLEKSAEYIILODFCHAIRKGVKYQQIIOGIQ   243
DB	121   VVNPRFKMSSTLNLVAGA PKSGI VWLEKSAEYIILODFCHAIRKGVKYQQIIOGIQ   180
QY	244   QLVKETGYTKTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYHDKVSDEAVNKIRLDIT   303
DB	181   QLVKETGYTKTPQGLFTPSPEIVKYTHKLAMERYAVFTDYHDKVSDEAVNKRILDIT   240
QY	304   EEOLEKEKPPEADPYEIIISFNVNAKEVRSIYLMNEYKACDGDLTSLRNVSCEVMFKTL   363
DB	241   EEOLEKEKPPEADPYEIIISFNVNAKEVRSIYLMNEYKACDGDLTSLRNVSCEVMFKTL   300
QY	364   HGSALFQGGQOVLCTVPFDSLSEGIKSDOYTAINGIKDKNFMHYEPFPATNEIGKV   423
DB	301   HGSALFQGGQOVLCTVPFDSLSEGISDOYTAINGIKDNFMHYEPFPATNEIGKV   360
QY	424   TGLNRRELGHGALAEKALYPVI PRDPFTTIRTVSEVLSENGSSSNASA CGSIALMDSGV   483
DB	361   TGLNRRELGHGALAEKALYPVI PRDPFTTIRTVSEVLSENGSSSNASA CGSIALMDSGV   420
QY	484   PISSAVAGVAIGLVTKTDPEKEGEIDYLLDILGIEDYNGDMDKIAGTKGTITALOAD   543
DB	421   PISSAVAGVAIGLVTKTDPEKEGEIEDYLLDILGIEBVGMDPKIAGTKGTITALOAD   480
QY	544   IKLPIPIKIIMEA IQOASVAKKELIQIMNTKISRPASRKENGVEVTVQVPLSKRAKF   603
DB	481   IKLPIPIKIIMEA IQOASVAKKELIQIMNTKISRPASRKENGVEV-----   527
QY	604   VGPGGYNNKKLQALETGVTISQVDEETFSVFAPTPSVMEHARDFITEICKDQEQOLFEGA   663
DB	528   -----GVTSQVDEETFSVFAPTPSVMEHARDFITEICKDQEQOLFEGA   572
QY	664   VYTATITEIRD TGWVKLYPNMTAVLHNTOJLNERL   700
DB	573   VYTATITEIRD TGWVKLYPNMTAVLHNTOJLDERKI   609
RESULT 5	
ID	ADC39092
XX	ADC39092 standard; protein; 540 AA.
AC	ADC39092;
DT	18-DEC-2003 (First entry)
XX	

DE Novel human NOVX polypeptide SEQ ID NO: 34.  
 XX anti-diabetic; cytoprotective; immunomodulator; anorectic; antilipemic;  
 KM noctropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;  
 KM antiasthmatic; antinflammatory; hypotensive; antiarteriosclerotic;  
 KM hemostatic; osteopathic; gene therapy.; NOVX; diabetes; obesity; cancer;  
 KM lymphoma; uterine cancer; prostate cancer; dyslipidemia; anorexia;  
 KM wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;  
 KM cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;  
 KM hyperextension; atherosclerosis; hemophilia; graft-versus-host disease;  
 KM Alldright hereditary osteodystrophy.  
 XX Homo sapiens.  
 XX WO2003010327-A2.  
 PD 06-FEB-2003.  
 XX 02-MAY-2002; 2002WO-US014199.  
 XX 02-MAY-2001; 2001US-0288063P.  
 PR 03-MAY-2001; 2001US-0288395P.  
 PR 07-MAY-2001; 2001US-0289087P.  
 PR 09-MAY-2001; 2001US-0289817P.  
 PR 09-MAY-2001; 2001US-0289818P.  
 PR 11-MAY-2001; 2001US-0290194P.  
 PR 14-MAY-2001; 2001US-0290753P.  
 PR 15-MAY-2001; 2001US-0291181P.  
 PR 16-MAY-2001; 2001US-0291243P.  
 PR 18-MAY-2001; 2001US-0292001P.  
 PR 21-MAY-2001; 2001US-0292374P.  
 PR 22-MAY-2001; 2001US-0292587P.  
 PR 23-MAY-2001; 2001US-0293107P.  
 PR 25-MAY-2001; 2001US-0293747P.  
 PR 29-MAY-2001; 2001US-0294109P.  
 PR 29-MAY-2001; 2001US-0294110P.  
 PR 30-MAY-2001; 2001US-0294434P.  
 PR 31-MAY-2001; 2001US-0294827P.  
 PR 12-JUL-2001; 2001US-0304879P.  
 PR 31-JUL-2001; 2001US-0308901P.  
 PR 14-AUG-2001; 2001US-0312270P.  
 PR 17-AUG-2001; 2001US-0313416P.  
 PR 10-SEP-2001; 2001US-0318463P.  
 PR 27-SEP-2001; 2001US-0325683P.  
 PR 18-OCT-2001; 2001US-0330292P.  
 PR 28-NOV-2001; 2001US-0333873P.  
 PR 03-DEC-2001; 2001US-0336909P.  
 PR 03-DEC-2001; 2001US-0337552P.  
 PR 21-FEB-2002; 2002US-0359245P.  
 PR 01-MAY-2002; 2002US-00136826.  
 XX (CURA-) CURAGEN CORP.  
 PA  
 PI Muller CB, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;  
 PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zearhusen BD;  
 PI Paturajan M, Anderson DW, Mezes PS, Peyman JA, MacDougall JR;  
 PI Padigaru M, Raseelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;  
 PI Edinger SR, Ellerman K;  
 XX  
 XX WPI; 2003-239445/23.  
 DR N-PSDB; ADC39091.  
 XX  
 XX New NOVX polypeptides and polynucleotides, useful in gene therapy,  
 PT particularly for treating or preventing a syndrome associated with a  
 PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,  
 PT hypertension or hemophilia.  
 XX  
 XX Claim 1; SEQ ID NO 34; 748bp; English.  
 XX  
 CC The invention relates to new isolated NOVX polypeptides, the genes  
 CC encoding them or sequences having at least 95% identity to the amino acid  
 CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,  
 CC particularly in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, which includes a pathology associated  
 CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for  
 CC treating, preventing or alleviating pathology associated with NOVX  
 CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and  
 CC polypeptide are especially useful for treating or preventing e.g.  
 CC diabetes, obesity, cancers (e.g. lymphoma, uterine cancer or prostate  
 CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,  
 CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's  
 CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,  
 CC graft-versus-host disease or Alldright hereditary osteodystrophy. The DNA  
 CC encoding the protein is useful in gene therapy for treating the above  
 CC conditions. These are also useful in developing powerful assay system for  
 CC functional analysis of various human disorders, as well as in diagnostic  
 CC applications. This sequence represents one of the NOVX proteins of the  
 CC invention.  
 CC  
 CC Sequence 540 AA;  
 XX  
 XX Query Match 64.4%; Score 2289; DB 7; Length 540;  
 XX Best Local Similarity 86.8%; Pred. No. 9e-191;  
 XX Matches 455; Conservative 22; Mismatches 47; Indels 0; Gaps 0;  
 QY 1 DGPFLPRRRALTOLOVRALMSAGRAVAVDIGNRLAISGKRLARPADGSAVVGSGD 60  
 DB 17 DGPPCLPGWDRALTOLOVRALSSGSAVAMDMGNKRLBISGKRLARFANGSAVVGSGD 76  
 QY 61 TAVWTVASKTPSPSPQMPVVDYRQAAAGRIPTNYLAREGTSDEKELTGRIDRS 120  
 DB 77 TAVMATASKTPSPSPQMPVVDYRQAAAGRIPTNYLAREGTSDEKELTGRIDRS 136  
 QY 121 IRPLFPAGPYDTQVLCNLAVDQVNEPDVLAINGASVALSLSDIPNNGPVGAVRIGIID 180  
 DB 137 VRIFPAGPYDTQVLCNLAVDQVNEPDVLAINGASVALSLSDIPNNGPVGAVRIGIID 196  
 QY 181 GEVYNPFRKEMSSSTNLVYAGPKQIYMLASANILOQDFCHAIKVCYKTTQOIIQ 240  
 DB 197 GECVNNPFRKEMSSSTNLVYAGPKQIYMLASANILOQDFCHAIKVCYKTTQOIIQ 256  
 QY 241 GIQQLVKEIGVTKTPQKLPSPSEIVKYTHKLMERLYAVFTYEHDKVSRDAVVKIR 300  
 DB 257 GIQQLVKEIGVTKTPQKLPSPSEIVKYTHKLMERLYAVFTYEHDKVSRDAVVKIR 316  
 QY 301 LDTEQLKEKEPEADPYEIIIESFNVVAKEVFRSIVLNEYKRCDSRDLTSLNVSCEVDMF 360  
 DB 317 LDTEQLKEKEPEADPYEIIIESFNVVAKEVFRSIVLNEYKRCDSRDLTSLNVSCEVDMF 376  
 QY 361 KTLHGSAIPFORQTOVLCCTVPDSLESIGKSDOYITAINIGIKDNFMILHFEPPYATNEI 420  
 DB 377 KTLHGSAIPFORQTOVLCCTVPDSLESIGKSDOYITAINIGIKDNFMILHFEPPYATNEI 436  
 QY 421 GKVTGLNRRELGHGALAEKALYPIPRDPPTTRVTSSEVSSNGSSMAACGGSLLAMD 480  
 DB 437 VKVTGINRRRELGHGALAEKALYPIPRDPPTTRVTSSEVSSNGSSMAACGGSLLAMD 496  
 QY 481 SGVPISSAVGVALGVLTKTPDEKGEIEDYRLLDIIGIEDYNG 524  
 DB 497 SGVPISSAVGVALGVLTKTPDEKGEIEDYRLLDIIGIEDYNG 540  
 XX  
 XX RESULT 6  
 XX AAB92684  
 ID AAB92684 standard; protein; 504 AA.  
 AC AAB92684;  
 XX 26-JUN-2001 (first entry)  
 DT  
 XX Human protein sequence SEQ ID NO:11065.  
 DE  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KM  
 XX Homo sapiens.  
 OS  
 XX



CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.

XX Sequence 504 AA:

Query Match 64.2%; Score 2285; DB 7; Length 504;

Best Local Similarity 100.0%; Pred. No. 1.8e-190;

Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 DGPFLRRRRALTOLOVRALMSGSRVAVDIGNRKLKLEISSGKLARFADGSAVVOGSD 60
DB 18 DGPFLRRRRALTOLOVRALMSGSRVAVDIGNRKLKLEISSGKLARFADGSAVVOGSD 77
QY 61 TAVVAVTSKTPSPQFMPLVAVYRQKAAAGRIPTNYLREVGTSDEKILTSRIIDRS 120
DB 78 TAVVAVTSKTPSPQFMPLVAVYRQKAAAGRIPTNYLREVGTSDEKILTSRIIDRS 137
QY 121 IRPLFPAGFYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 180
DB 138 IRPLFPAGFYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIID 197
QY 181 GEYVAVPTKREMSSTLNLVVAGAPKSOIWMLEASENILLODPCHAIRKGVKTTQOIIO 240
DB 198 GEYVAVPTKREMSSTLNLVVAGAPKSOIWMLEASENILLODPCHAIRKGVKTTQOIIO 257
QY 241 GIOOLVKEGTGKRTPOKLTSPSEIYKTHKLMERLYAVFTDYEHDKSRDAVANKIR 300
DB 258 GIOOLVKEGTGKRTPOKLTSPSEIYKTHKLMERLYAVFTDYEHDKSRDAVANKIR 317
QY 301 LDTEBQKEKRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMF 360
DB 318 LDTEBQKEKRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMF 377
QY 361 KTLHGSLFPQRCQOVCTTTFDSLESGIKSDQVITAINIKDKNFMHYEPYATNEI 420
DB 378 KTLHGSLFPQRCQOVCTTTFDSLESGIKSDQVITAINIKDKNFMHYEPYATNEI 437
QY 421 GKVTGLNRRELHGALAEKALYPYIPRDPF 450
DB 438 GKVTGLNRRELHGALAEKALYPYIPRDPF 467

```

# RESULT 8

ABG17275 standard; proteoin: 899 AA.

XX ABG17275;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #17266.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Dermanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS81462.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20; SEQ ID NO 47634; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantifying a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic  
XX amino acid sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pcr\_sequences

XX Sequence 899 AA;

Query Match 55.7%; Score 1982; DB 4; Length 899;

Best Local Similarity 73.2%; Pred. No. 1.5e-163;

Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

```

QY 10 DRALTOLOVRALMSGSRVAVDIGNRKLKLEISSGKLARFADGSAVVOGSDTAVMTAVS 69
DB 440 DRALTOLOVRALMSGSRVAVDIGNRKLKLEISSGKLARFADGSAVVOGSDTAVMTAVS 499
QY 70 KTRPSQFMPLVAVYRQKAAAGRIPTNYLREVGTSDEKILTSRIIDRSIRLPFAGY 129
DB 500 KTRPSQFMPLVAVYRQKAAAGRIPTNYLREVGTSDEKILTSRIIDRSIRLPFAGY 521
QY 130 FYDTQVLCNLAVDGVNEBDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEYVAVNPT 189
DB 522 -----VGMDGECVAVPTR 535
QY 190 KEMSSSTLNLVVAGAPKSOIWMLEASENILLODPCHAIRKGVKTTQOIIOGIVKVKET 249
DB 536 KEMSSSTLNLVVAGAPKSOIWMLEASENILLODPCHAIRKGVKTTQOIIOGIVKVKET 595
QY 250 GVTKRTPOKLTSPSEIYKTHKLMERLYAVFTDYEHDKSRDAVANKIRLDTBEOLKE 309
DB 596 GVTKRTPOKLTSPSEIYKTHKLMERLYAVFTDYEHDKSRDAVANKIRLDTBEOLKE 655
QY 310 KRPADPYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMFKTLHGSLF 369
DB 656 IPEVVDLYEIIIESFNVAKEVRSIVLNEYKRCDDRDLTSLRNVSCEVDMFKTLHGSLF 715
QY 370 QRGOTQVLCVTPDPSLESGIKSDQVITAINIKDKNFMHYEPYATNEIIGVTGLNRR 429
DB 716 QRGOTQVLCVTPDPSLESGIKSDQVITAINIKDKNFMHYEPYATNEIIGVTGLNRR 775
QY 430 ELHGALAEKALYPYIPRDPFPIRVTSEVLESGSSMASACGGSIALMDSGVPISAV 489
DB 776 ELHGALAEKALYPYIPRDPFPIRVTSEVLESGSSMASACGGSIALMDSGVPISAV 818
QY 490 AGVAGLVATKDEPKGIEDYRLLTDIAGIEDNGMDDFIAGTNKGTITLQADIKLPGI 549
DB 819 GVMAGLATKTDLEKGEIEDYRLLTDIAGIEDNGMDDFIAGTNKGTITLQADIKLPGI 877
QY 550 PIKIUMEAIQOASVAKK 566

```

Db 878 TMKIWEAIQOASVAKR 894

RESULT 9  
ABG08546  
ID ABG08546 standard; protein; 899 AA.  
XX  
AC ABG08546;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #8537.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00549217.  
PR 23-AUG-2000; 2000US-00549167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS72733.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 38905; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 899 AA;

Query Match 55.7%; Score 1982; DB 4; Length 899;  
Best Local Similarity 73.2%; Pred. No. 1,5e-163;  
Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

OY 10 DRALTOLOVRALMSAGSRAVAVDLGNRKLEISSGKLARFADGSVAVVSGDTAVMTAVS 69  
DB 440 DRALIQLQVRLALSSTSQSAVAMDIGNKKLEISSGKLARFAGSAVAVVSGDTAVMTAVS 499

OY 70 KTKPSPSQFMPLVVDYRQKAAAGRIPTNYLRREVGTSDEKIELTSRIIDSIRPLPPAGY 129  
DB 500 KTKPSPSQFMPLVVDYRQKAA----- 521  
OY 130 FYDTQVLCNLAVDGVNPEVDVLAINGASVALSLSDIPWNGPVGAVRIGITIDGEVYVNPTR 189  
DB 522 -----VGMTDGEVCVNPTR 535  
OY 190 KEMSSSTLNLVVGAPKRSQIVMLEASAEINLQDPCHAIKVGKYYTQOIIOGIQOLVKET 249  
DB 536 KEMSSSTLNLVVGAPKRSQIVMLEASAEINLQDPCHAIKVGKYYTQOIIOGIQOLVKET 595  
OY 250 GYTKRTPQKLFPTSPPEIVKTHKLAMERLYAVPTVDYHDVSDDEAVNKRIRLDTREOLKE 309  
DB 596 GYTKRTPQKLFPTSPPEIVKTHKLAMERLYAVPTVDYHDVSDDEAVNKRIRLDTREOLKE 655  
OY 310 KPPEADPEYIIIESFNVAKEVFRSIVLEKRCDCGDLTSLRVNCSCEVDMFKTLHGSALE 369  
DB 656 IPEVDLYEIIIESFNVAKEVFRSIIINEYKRCDCGDLTSLRVNCSCEVDMFKTLHGSELE 715  
OY 370 QRGQTVLCTVTPDSLESGIKSDQVITAINGIKDKPMHYEPPPYATNEIGKVTGLNRR 429  
DB 716 QRGQTVLCAVTPDSLESGIKDLRVITTINGIKDKPMHYEPPPYATNEIVKVTGINRR 775  
OY 430 ELHGALAEKALPVIPRODPFTIRVTSBYLSENGSSSMASACGSLALMDSGVPISSAV 489  
DB 776 ELGPGALAEKALPVIPRODPFTIRVTSBYLSENGSSSMASAC----- 818  
OY 490 AGVAILGVTKTDEKEIEDYRLLTDLIGIEDNGDMDFKIAGTNKGITLALQADIKLPGI 549  
DB 819 -GVAMGLATKTIDLEKGEIEDYRLLTDLIGIEAINGDMDFKIAGTNKGITLALQADIKLPGI 877  
OY 550 PIKIWEAIQOASVAKK 566  
DB 878 TMKIWEAIQOASVAKR 894

RESULT 10  
ADE08997  
ID ADE08997 standard; protein; 899 AA.  
XX  
AC ADE08997;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Novel protein-related contig polypeptide sequence #63.  
XX  
KW novel gene; novel protein; tissue marker; molecular weight marker;  
KW chromosome marker; genetic disorder; contig.  
XX  
OS unidentified.  
XX  
PN W02003054152-A2.  
XX  
PD 03-JUL-2003.  
XX  
PD 10-DEC-2002; 2002WO-US039555.  
XX  
PF 10-DEC-2001; 2001US-0339739P.  
XX  
PR 11-DEC-2001; 2001US-0339453P.  
PR 14-MAR-2002; 2002US-0365091P.  
PR 14-MAR-2002; 2002US-0365384P.  
PR 12-APR-2002; 2002US-0372381P.  
PR 12-APR-2002; 2002US-0372615P.  
PR 22-APR-2002; 2002US-00128558.  
PR 24-APR-2002; 2002US-0376045P.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xie AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

DR WPI; 2003-569235/53

PT New polynucleotides, useful for expressing recombinant proteins for  
PT analysis, characterization or therapeutic use, or as markers for tissues  
PT in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2541; 1177pp; English.

CC The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence was used in the  
CC exemplification of the invention

**SQ Sequence 899 AA;**

Query Match	Score	1982; DB 7;	Length
55.74;	899;		

Matches 408; Conservative 18; Mismatches 29; Indels 102; Gaps 2;

QY	10	DRALPOLQVRALMSKSGSFAVVDLGNRLKLEISSGGLAFADQSAVQSSDITVMTAVS	69
Db	440	DRALPOLQVRALLSTSGSAVAMDGNRLKLEISSGGLAFANCSAVIQSDITVMTAVS	459
QY	70	KTTPSPSPQFPLVVVDYRQKAAAGRLPTMYLRREVGTSDEILTSIHDSIRPLPAGY	129
Db	500	KTTPSPSPQFPLVVVDYRQKAAA-----	521
QY	130	FYDTQVLNLLAVDGVNEPDVLAINGASVALSLSDIPMNGPVGAVARIGITDEYVNPTR	189
Db	532	-----VGMTDGCYVNPTR	535

Qy	190	KEMSSSTNLVVAAGPKSIOIWLLEASAEINLQDDPCHAIKVGVKTYQOIIQGIQOLVKET	2.49
		:	
Db	536	KEMSSSTNLVVAAGPKSQIWLLEASAEINLQDDPCHAIKVGVKTYQOIIQGIQOLVKET	59.95
		:	
Qy	250	GVTKSTPKLFTPSPEIVCYTKHKLAMERIYAAPVDYEHDKVSDDEAVNKIRLDTBEOLKE	3.09
		:	
Db	596	GVTKSTPKLFTPSPEIVGHAHKLIMERIYAAPVDYEHDKISDEAVNKIRLDTBEOLKE	65.55
		:	
Qy	310	KPEADPEYIIIESFNVAKEVFRSIYLTNEYKRCDSGLDLSLRVSCSEVDMAFKTLHGSALF	3.69
		:	
Db	656	IPFBDVLRIIISFNTVAKEVFRSIIITNEYKRCDSGLDLSLRRIISCEVDMAFKTLHGSALF	71.51

Qy	Dy
370	370
RGCGQVYCTVAFSDLSSEGIKSDQVTTANGIDKDKPMALHYEPPPATNIGKATYGLNRR	RGCGQVYCTVAFSDLSSEGIKSDQVTTANGIDKDKPMALHYEPPPATNIGKATYGLNRR
716	716
QRGQRLCAVAFDLSSESIKLDRIVTITTINGIDKDKPMALHYEPPPATNINIVATYGINRR	QRGQRLCAVAFDLSSESIKLDRIVTITTINGIDKDKPMALHYEPPPATNINIVATYGINRR
430	430
ELGHGALAEKALYPIVIRPDPFTIRTSYVLENSGSSSMASACCGSLALMDSGVPISSAV	ELGHGALAEKALYPIVIRPDPFTIRTSYVLENSGSSSMASACCGSLALMDSGVPISSAV
776	776
ELGGPALAEKALYPIVIRPDPFTIRTSYVLENSGSSSMASAC-----	ELGGPALAEKALYPIVIRPDPFTIRTSYVLENSGSSSMASAC-----
	818

[illegible]

RESULT 11  
ADE07933  
ID ADE07933 standard; protein; 476 AA.

AC	ADE07933;	
XX		
DT	29-JAN-2004	(first entry)
XX		
DE	Novel protein (useful for identifying genetic disorders)	#88

XX novel gene; novel protein; tissue marker; molecular weight marker;  
KM chromosome marker; genetic disorder.  
KM

OS Unidentified.  
YY

PN WO2003054152-A2

PD 03-JUL-2003

PF 10-DEC-2002; 2002WO-US039555.

PR 10-DEC-2001; 2001US-0339739P.  
PR 11-DEC-2001; 2001US-0339452P.

PR 14-MAR-2002; 2002US-0365091P.  
PR 14-MAR-2002; 2002US-0365091P.

PR 12-APR-2002; 2002US-0372381P.  
PR 13-APR-2003; 2003HE-037261EP.

PR 22-APR-2002; 2002US-00128558.  
PR 24-APR-2003; 2003US-037604EP

XX  
XX (HYGE) HYGEN INC[illegible]

PI Ghosh M, Xue AJ, Wehrman T,

XX  
WPB - 2003-EC0336/EC3

DR N-PSDB; ADE07022.  
XX

PT New polynucleotides, useful for expressing recombinant proteins for analysis, characterization or therapeutic use. Or as markers for trans-

PT in which the corresponding protein is preferentially expressed.

PS Claim 20; SEQ ID NO 999; 1177pp; English  
xx

CC The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially expressed, as molecular weight markers on gels, as chromosome

CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.

SQ Sequence 476 AA;

Query Match	Score	IDB #	Length
55.2%	1964	DB 7	476
70.1%	Dred No.	30-153	

Matches 408; Conservative 16; Mismatches 30; Indels 112; Gaps 4

QY	1	DGPFLLPRDRALTOLOVALMSSAGSRVAADLGRKLEISSGKLARFADSSAVVQSG	66
Db	18	DGPFLLPRDRALTOLOVALMSSAGSRVAADLGRNRP	67
		-----ALF-----FMFRSGD	
QY	61	TAVMTAASKTSPSPQFMPLVVDPRQKAAAGRIPTNYLRREVGTSDKEILTSRIIDRS	120
Db	68	TAVMTAASKTSPSPQFMPLVVDPRQKAA	98
		-----	
QY	121	IRPLFPAGFYDTQVLCHLLAVDGVNEBPVLAINGASVALSLSDIPIMNGPVGAVRIGIID	180
Db	99	-----VGMTD	103

QY	181	GEYVNPPIRKEMSSSTLNVAGAPKSCQVYMLASANILODFCAIKVGVKXTOOIIQ	240
Db	104	GECVNPPIRKEMSSSTLNVAGAPKSCQVYMLASANILODFCAIKVGVKXTOOIIQ	163
QY	241	GIQQLVKETGVTKRPPQKLFTRPSPIVYKTHGLAMERLYAVFTDYEHDVSRDEAVNKR	300
Db	164	GIQQLVKETGVTKRPPQKLFTRPSOEIVKFAHKLIMERLYAVFTDYEHDKISRDEAVNKR	223
QY	301	LDTEQLTEKPEADPYEIIISFNVYAKVEPRSIYLNKYRCDCGRDLYLSRNVSCVDMP	360

Db 224 LDTEBQKEKFEPEVDLYEIIIESFNTVAKEVFRSIIINEYKRCDCGRDLTLRNISCEVDMF 283  
Qy 361 KTLHGSALFQRGQTQVLCYVTFDSLESIGKSDQVITAINGIKDXNFMHLHYEPFYATNEI 420  
Db 284 KTLHGSSELPQRGQTQVLCYVTFDSLESIGKSDQVITTINGIKDXNFMHLHYEPFYATNEI 343  
Qy 421 GKVTGLNRRELGHGALAEKALYPIPRDPFPTIRVTSEVLESNSSSMASACGSLALMD 480  
Db 344 VKVTGIRNRRELGPALAEKALYPIPRDPFPTIRVTSEVLESNSSSMASAC----- 395  
Qy 481 SGVPISSANVAGVLTCTDPEKGEIEDYRLTDIIGIEDYNGMDMPKIKGTNGKGTAL 540  
Db 396 -----GVANGLATKTDLEKGEIEDYRLTDIIGIEAYNGMDMPKIKGTNGKGTAL 445  
Qy 541 QADIKLPQIPKIKIYWEAIIQASVAKK 566  
Db 446 QADIKLPQITWKIKIYWEAIIQASVAKR 471

RESULT 12  
ABG17276  
ID ABG17276 standard; protein; 439 AA.  
XX  
AC ABG17276;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17267.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Dmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
DR N-PSDB; AAS81463.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 47635; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 439 AA;  
Query Match 53.9%; Score 1917; DB 4; Length 439;  
Best Local Similarity 78.7%; Pred. No. 2.2e-158;  
Matches 396; Conservative 6; Mismatches 15; Indels 86; Gaps 3;  
Qy 1 DGPFLPRRRALTOLOVRALWSSAGRAVAVDGNRKLETSSGKLRFADGSAVVGSD 60  
Db 18 DGPFLPRRRALTOLOVRALWSSAGRAVAVDGNRKLETSSGKLRFADGSAVVGSD 77  
Qy 61 TAVVNTAVSKTKPSQSPQFPLVVDYRQKAAGRIPTNYLRREVGTSDKEILTSRIDRS 120  
Db 78 TTVVNTAVSKTKPSQSPQFPLVVDYRQKAA----- 108  
Qy 121 IRPLFPAGYFYDQVLCNLLAVDGVNEPDLAINGASVALSLSDIPMNGPVGAVRIGIID 180  
Db 109 -----AVRIGIID 116  
Qy 181 GEYVYNPTRKEMSSSTINLVVAGAPKQIYMLEASAENIIQDPFGHAKYGVYKTOQIIQ 240  
Db 117 GEYVYNPTRKEMSSSTINLVVAGAPKQIYMLEASAENIIQDPFGHAKYGVYKTOQIIQ 176  
Qy 241 GIQQLVETGVTKRTPQKLFTPSPETVYTHKLAERLYAVFTDYEDHKYRDEAVVKIR 300  
Db 177 GIQQLVETGVTKRTPQKLFTPSPETVYTHKLAERLYAVFTDYEDHKYRDEAVVKIR 236  
Qy 301 LDTEBQKEKFEPEADPYEIIIESFNVVAKEVFRSIVINEYKRCDCGRDITSLRNISCEVDMF 360  
Db 237 LDTEBQKEKFEPEADPYEIIIESFNVVAKEVFRSIVINEYKRCDCGRDITSLRNISCEVDMF 296  
Qy 361 KTLHGSALFQRGQTQVLCYVTFDSLESIGKSDQVITAINGIKDXNFMHLHYEPFYATNEI 420  
Db 297 KTLHGSALFQRGQTQVLCYVTFDSLESIGKSDQVITAINGIKDXNFMHLHYEPFYATNEI 356  
Qy 421 GKVTGLNRRELGHGALAEKALYPIPRDPFPTIRVTSEVLESNSSSMASACGSLALMD 480  
Db 357 GKVTGLNRRELGHGALAEKALYPIPRDPFPTIRVTSEVLESNSSSMASACRKPFSI-- 414  
Qy 481 SGVPISSAVA---GVALGVTKT 500  
Db 415 NGFRGSNFICCCRPYKGLVTKT 437

RESULT 13  
ABBS8546  
ID ABBS8546 standard; protein; 748 AA.  
XX  
AC ABBS8546;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 2430.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
PD 23-MAR-2001; 2001WO-US009231.  
XX  
PF 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX



PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li FMD, Myers EW,  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB; ABL02649.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell  
PT interactions.

PS Disclosure; SEQ ID NO 2430; 21pp + Sequence Listing; English

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116157-AB143511), expressed DNA sequences (AB010840-AB116157) and the encoded proteins (AB557377-AB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from <http://www.ncbi.nlm.nih.gov/pubmed/patent>.

**SQ Sequence 748 AA;**

Query Match	51.8%	Score	1842.5	DB 4	Length	748			
Best Local Similarity	52.8%	Pred	No. 1.8e-151						
Matches	364	Conservative	124	Mismatches	170	Indels	31	Gaps	6

```

QY 1 RALMSAGSAVAVVD- --GNRKLEISGCKLARPDGSAVVOGQTAMVAWAKTSP 75
D 29 RGIQSSNGEABEVEVNFSGRMNTSSRLRPFANGTAVCOMDGTAMVTAWAKAPN 88
QY 76 SQ-FMPLVYDROKAAAGRIPTNYLRREVGTSDKEILTSRIIDSRINRLEPAGYFDQ 134
D 89 GQGFMPVLVYDRILKNASGRI PMNFMRRELGTSGKEILSARLIDRSLPLFHKDYRETQ 148
QY 135 VLCNLAVDGVNEPDLTANGASVALSLSDIPMNPVGAVRGIIIDGYYVNPTRKEMS 199
D 149 LVCMNLAMAVHSDPVLAINAASMLSLSDIPMNPIGAVRGVGLCDGVVLNPRRELQT 208
QY 195 STNLVAVAGPKSOI WLEASAEINIKODFCIAIKGVYVTOIIOGLOLWKEGTATK 255
D 209 SOLDLVVSATKONLVMLBEGKAVVUQDOLLKAIQGTREAOFIHEIRLOKAYGOKR 266
QY 255 TPQKLTPTSEBIYKYTHKLAMERLYAVFTDYEDHVRSDAEVANKRLDTBESQLEKPEA 314
D 269 EVEVAABVDDELCKAVRSMCEMLRREI PDSYTHDKMSRDNANAEVRSVUIKWSSPPT 328
QY 315 DPEIIESEFVAVAKVFRSLVANYEYRCGGRBLTSLBVSCEVDMFKTLHSAIPFGQT 374
D 329 EPLSLIEQFQTSRTIFRELIFERGLRCGRYDQIRNISCVDMYKPLHOSALPFGQT 388
QY 375 QVLCVTPDLSBSGISKSDQVITAIN--GIKDKNFMILHYEPFATNEIGKTYGLNREL 432
D 389 QVCTYSLSBQESAMKUDS-LAALBDSGKAKCNPLHTEFPYATGVEGRGCPVGREMG 444
QY 433 HGALAEKALVPVIRPDPPTIRVTSVLESNGSSSMASACGSLALMDGVPISAVAGV 492
D 448 HGALAEBSLPLTPVNDYPTFVRLTSSVLESNGSSSMASACGSLALMDAGVPISAPAGV 507
QY 493 AIGJAVTK--DPEKGELEDYRLTLDIGEDNAGMDPKIAGNNKITLMOQIDILPGPI 553
D 508 AIGJAVTKFEENDTKTHLODRILTLDIGIEDNYGMDH-----M 544
QY 552 KIWEALIOQASVAKKEILQIMNTKISKPASREKNGPVETQVYBLSRAKAVPGVNL 611
D 545 KYWVESLOKATDAKSNILIDIMSEAINBRPKYPRZESMPVSETULVBPQOAOIGPSGLH 604
QY 612 KKLQAEFGVITISOVDEBTSVVPKPTYSVNHKARDPTELCDQDQOOLEFQAVNTATITE 671
D 605 KRYLEBGTSLVAVDEHVFVBPQAAWDEKKEILBGWYMERVPDLEFGIYATKITE 664

```

```
Oy      672 IRDTGVWVLTYPNMTAVLHNTQLDNEKL 700C
          :||||| |:| ||||:::
Db      665 IRTDVWVILYPSMPALHNSQLDQRI 693
```

RESULT 14  
ABG08547  
ID ABG08547 standard; protein; 541 AA

AC ABG08547

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8538

KW Human; chromosome mapping; gene mapping; gene therapy; forensic food supplement; medical imaging; diagnostic; genetic disorder.

## OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US008631.

PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT,

DR WPI; 2001-639362/73  
DR N-PSDB; AAS72734.

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity.

PS Claim 20; SEQ ID NO 38906; 103bp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AB000010-AB030377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/publ/published\\_pct\\_sequences](http://wipo.int/publ/published_pct_sequences)

**SQ Sequence 541 AA;**

Query Match	Score	DB 4	Length
34.5%	1228.5	541	

Best Local Similarity 51.7%; Pred. No. 5.3e-98;  
Matches 306; Conservative 48; Mismatches 133; Indels 105; Gaps 18

QY 97 TNYLRREVGTSKIELTSRIIDRSIRPLFPAGYFYDTQVL CNLLAVDGVNEPDVLAINGA 156

```

Db      6 TNEIGKVTGLNRELGHGALAEKALYPIVPRDPTIRVTSEVLESNGMR----- 55
Qy      157 SVALSLSDIPWNGPVAVRIGITDGEVYVNPTRKEMSSSTNLVAVAGP-----KSGIV 210
Db      56 -----RIG-----QPPASSRPVRVRAPLPPLLEVRAPLP 86
Qy      211 MLEASANILOODFCHAIKVGKTYTOI--IQIGIOLVKEGVTKRTPOKLFPTSPELVX 268
Db      87 ASRPSGEVLCTVTF--DSLSEGIKSDQVITANGIK-----DNKF----- 124
Qy      269 YTHKLAERLYAVFTDYHDKVSDEAVNKIRLD---TEBOLKEKPEADPYEIIESFN 324
Db      125 -----MAHYEPFPYATNEIGKVTGLNRELGHGALAEKALYPIVPRDPTIR----- 171
Qy      325 VVAKEVPSIVLNEYKRCDSGLTS--LRNVSCE-----VDMFTLHGSALEFORQOTVL 377
Db      172 RVTSEVLES---NGMRRIQGPASSRPVRVRAPLPPLLEVRAPLPAS---RPSGEVL 224
Qy      378 CTVPFDSLESGIKSDQVITANGIKDKNFMHYEPFPYATNEIGKVTGLNRELGHGALA 437
Db      225 CTVPFDSLESGIKSDQVITANGIKDKNFMHYEPFPYATNEIGKVTGLNRELGHGALA 284
Qy      438 EKALYPIVPRDPTIRVTSEVLESNGSSMASACGSLALMDGVPISAVAG--VAIGL 496
Db      285 EKALYPIVPRDPTIRVTSEVLESNGSSMASACGSLALMDGVPISAVAGRTXDW 344
Qy      497 VTKTDPKEGEIEDYRLTLDI--LGIEDYNGDMFKIAGTNKIGITALQADIKLPGIPIKIVM 555
Db      345 SPKDDPEKGEIRIKIIVCKQIFLGIEDYNGDMFKIAGTNKIGITALQADIKLPGIPIKIVM 404
Qy      556 EAIQOASVAKKEIIQINMKITISKPRASKENGPPVETVQVLSKRAKVGCGYNLKKLQ 615
Db      405 EAIQOASVAKKEIIQINMKITISKPRASKENGPPVETVQVLSKRAKVGCGYNLKKLQ 464
Qy      616 AETGVTISQVDEET--FSVFAPTPSVMEHARDP--ITEIKDOQEOOL--EFGAV 664
Db      465 AETGVTISQVDEETVFCMLHQPVLMEGKRLSTETICKDOEOAIRNLGAV 516

RESULT 15
ABU24041
ID ABU24041 standard; protein: 703 AA.
AC ABU24041;
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #9568.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Clostridium acetobutylicum.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trivick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX DR N-PSDB; ACA27911.

```

```

XX      XX New antisense nucleic acids, useful for identifying proteins or screening
PT      PT for homologous nucleic acids required for cellular proliferation to
PT      PT isolate candidate molecules for rational drug discovery programs.
XX      PS Claim 25; SEQ ID NO 51965; 1766pp; English.
XX      CC The invention relates to an isolated nucleic acid comprising any one of
CC      CC the 6213 antisense sequences given in the specification where expression
CC      CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC      CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC      CC encoding a polypeptide whose expression is inhibited by the antisense
CC      CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC      CC polypeptide or its fragment whose expression is inhibited by the
CC      CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC      CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC      CC proliferation or the activity of a gene in an operon required for
CC      CC proliferation; (7) identifying a compound that influences the activity of
CC      CC the gene product or that has an activity against a biological pathway
CC      CC required for proliferation, or that inhibits cellular proliferation; (8)
CC      CC identifying a gene required for cellular proliferation or the biological
CC      CC pathway in which a proliferation-required gene or its gene product lies
CC      CC or a gene on which the test compound that inhibits proliferation of an
CC      CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC      CC compound's activity; (11) a culture comprising strains in which the gene
CC      CC product is overexpressed or underexpressed; (12) determining the extent
CC      CC to which each of the strains is present in a culture or collection of
CC      CC strains; or (13) identifying the target of a compound that inhibits the
CC      CC proliferation of an organism. The antisense nucleic acids are useful for
CC      CC identifying proteins or screening for homologous nucleic acids required
CC      CC for cellular proliferation to isolate candidate molecules for rational
CC      CC drug discovery programs, or for screening homologous nucleic acids
CC      CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC      CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC      CC the target prokaryotic essential genes. Note: The sequence data for this
CC      CC patent did not form part of the printed specification, but was obtained
CC      CC in electronic format directly from WIPO at
CC      CC ftp.wipo.int/pub/published_pct_sequences
XX      SQ Sequence 703 AA;
XX
XX      Query Match 33.9%; Score 1204.5; DB 6; Length 703;
XX      Best Local Similarity 39.6%; Pred. No. 1e-95;
XX      Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;
Qy      33 DLGNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTRSPS--QFMPLVVDYRQKAA 91
Db      8 DIAGRKLVKCGKTMGMSNCAMFISYGDITVMMNVNASEKRGRGIDPFPLISYBERQYS 67
Qy      92 AGRIPNYLRREVGTSDKEILTRIIDRSIRPLPAGYFYDTGYLQNLAVDGVNEPDVL 151
Db      68 VGKIPGCFVGRERPSSEKSLIHARAIDRPLRPLPKGYRNDVGVVCSVEDDNIPFIL 127
Qy      152 AINGASVALSLSDIPNMGPVAVRIGITDGEVYVNPTRKEMSSSTNLVAVAGPKSQIWM 211
Db      128 AMNGASVALCLSDIPPTVPATVSVGIDSKFVLPPLERREKSSLDLVC--ATNEVMM 186
Qy      212 LLEASANILOODFCHAIKVGKTYTOIIGIOLVKEGVTKRTPOKLFPTSPELVYTH 271
Db      187 LEAGADEIPEDMIAIDDEFNACQDPIVAFQEKAMEPGEKXTPE--LYHPKEIEIDVT 245
Qy      272 KLAERLYAVFTDYHDKVSDEAVNKIRLDTEBOLKEKPEADPYEIIESFNVAKEVF 331
Db      246 EFAPESIKEIM--YIDRDERNRNLRREIKESINSEPAEKYPD--DGADIDEVVYTLQKQV 302
Qy      332 RSVIINEYKRCDSGLTSLRNVSCEVDMFTLHGSALEFORQOTVCTVPFDSLESGIKS 391
Db      303 RNMLKEHRRPDRRDELRISCVDLLPRTHGSGLFTRGTLQVWMTVTTL---GPITG 358
Qy      392 DOVITANGIKDKNFMHYEPFPYATNEIGKVTGLNRELGHGALAEKALYPIVPRD 449
Db      359 AQVITDGLGVESKRYVHHNIFPPYSTGCVKPLRGPNRREIGHGALAEKALYPIVPRD 418

```

Qy 450 PFTIRVTSEVLSENGSSSMASACGSLALMDGVPISSAVAGVAIGLVTKTDPKGEIED 509  
 Db 419 PYTIRLVSEVLSSNGSTSQASVCGSTLALMDAGVPIKRPAGIAMGLITSEDLKBAV-- 476  
 Qy 510 YRLITDLIGIEDVNGDMDFKIAGTNKGITLQADIKLPGLPIKIWEALIOQASVAKKEIL 569  
 Db 477 ---ITDIOGLEDFPGDMDFKAGTEKGITAIQVDTKIHGLSKYCICKTAINDARKARLFI 533  
 Qy 570 QIMMKTSKPRASRKENGVPVETVQVPLSKRAKFFVPGGVNLKKLOAETGVTSQVDEET 629  
 Db 534 EKWACINEPKKEISTAPRAYTINIDTDKIRTLIGTGKTINKIIEBTGVKIDIREDET 593  
 Qy 630 FSVFAPTPSVWHBARDFITEICKDOEQLEFGAVYATITEIRDTGVWVLYPNMTAVL 689  
 Db 594 VFVLSSDADSANRALKMIIDDLTKD-----VKGGEVILGKVTKITNFGAFVEVLPGKEG-L 647  
 Qy 690 LAHTOLDNERLN 701  
 Db 648 WHISKLDINKVN 659

Search completed: January 28, 2005, 19:31:24  
 Job time : 170 secs

This Page Blank (Just for)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:18:21 ; Search time 44 Seconds  
(without alignments)  
1541.655 Million cell updates/sec

Title: US-09-907-907a-42

Sequence: 1 DGFPLPRRDRALTQLQVRA.....TAVLHNTQLDNERLTLLP 705

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1365.5	38.4	991	2 T48631	polynucleotide pho
2	1204.5	33.9	703	2 B97123	polyribonucleotide
3	1190.5	33.5	698	2 B89901	polyribonucleotide
4	1166.5	32.8	713	2 A82586	polyribonucleotide
5	1166.5	32.8	713	2 F97368	polyribonucleotide
6	1161	32.6	711	2 AD0901	polynucleotide pho
7	1158.5	32.6	749	2 G97782	hypothetical prote
8	1158	32.6	734	2 H85979	polynucleotide pho
9	1158	32.6	734	2 B91134	polynucleotide pho
10	1156	32.5	810	2 G75320	polynucleotide pho
11	1153	32.4	718	2 S74509	polynucleotide
12	1152	32.4	734	2 H65106	polyribonucleotide
13	1150	32.3	714	2 AC3497	polyribonucleotide
14	1146	32.2	705	2 AC0424	polyribonucleotide
15	1145	32.2	704	2 G83950	polynucleotide pho
16	1140.5	32.1	745	2 B71654	polyribonucleotide
17	1140.5	32.1	775	2 B70320	polyribonucleotide
18	1138.5	32.0	718	2 AD2355	polyribonucleotide
19	1137	32.0	719	2 P82831	polynucleotide pho
20	1134	31.9	709	2 B64056	polyribonucleotide
21	1132	31.8	705	2 S70691	polyribonucleotide
22	1130	31.8	723	2 AG1603	polynucleotide pho
23	1127	31.7	723	2 AC1241	polynucleotide pho
24	1121	31.5	709	2 B82298	polyribonucleotide
25	1115	31.3	707	2 B84973	polyribonucleotide
26	1114.5	31.3	694	2 B86615	polyribonucleotide
27	1114.5	31.3	694	2 D82009	polyribonucleotide
28	1112	31.3	707	2 C81161	polyribonucleotide
29	1111	31.2	706	2 H81943	probable polyribon

30	1106	31.1	709	2 S38883	polyribonucleotide
31	1105	31.1	712	2 B87253	polyribonucleotide
32	1097	30.8	701	2 C83052	polyribonucleotide
33	1091	30.7	708	2 A72264	polynucleotide pho
34	1068.5	30.0	695	2 E71463	probable polyribon
35	1067.5	30.0	722	2 D70200	polyribonucleotide
36	1066	30.0	702	2 C71269	probable polyribon
37	1064	29.9	693	2 G81725	polyribonucleotide
38	1063.5	29.9	773	2 G86856	hypothetical prote
39	1054	29.6	737	2 D95068	polyribonucleotide
40	1053	29.6	775	2 D97936	polyribonucleotide
41	1029.5	28.9	719	2 G81332	polyribonucleotide
42	1016.5	28.6	897	2 T06540	polyribonucleotide
43	1005.5	28.3	739	2 T10932	polyribonucleotide
44	1000.5	28.1	777	2 T49900	probable polyribon
45	998.5	28.1	773	2 H87015	hypothetical prote

## ALIGNMENTS

## RESULT 1

T48631  
polynucleotide phosphorylase - Arabidopsis thaliana  
N/Alternate names: protein T15N1.70  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C/Accession: T48631  
R/Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le  
submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z24493  
A/Accession: T48631  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-991 <BEV>  
A/Cross-references: UNIPROT:Q967G6; EMBL:AL163792  
A/Experimental source: cultivar Columbia; BAC clone T15N1  
C/Genetics:  
A/Map position: 5  
A/Intons: 107/3; 159/3; 216/3; 258/3; 312/3; 352/3; 398/3; 453/1; 495/1; 538/3; 612/1;  
A/Note: T15N1.70  
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match	38.4%	Score 1365.5	DB 2	Length 991
Best Local Similarity	41.7%	Pred. No. 5.4e-77		
Matches	287	Conservative 138	Mismatches 231	Indels 33
Gaps	11			
QY	24	SAGSRAY----	AVDGNRKLEISSGKLARPADSAAVQSGDTAVMTAVSKTRPSQF	78
DB	44	SAGTKILSFKEBEVSGSRVVSFETGKIARFANGSVVLGMDETIVLSTVCAKTDSPDF	103	
QY	79	MLVVDYRQKAAAGRIPTNYLRRVGTSDKEIITSRIIDRSIRPLPAGFYTOVLGN	138	
DB	104	LPLTVDOEKQYAOGLIPNTYMRREGAPKEBELLCGRILIRPIPLPPTGTVHVOIMAS	163	
QY	139	LIAVGVNEPVPVLAINGASVALISDIIPMGVPGAVGIIIGDGVVNPPEKEMSSSTLN	198	
DB	164	VLSDDGKDPDITLAANSSAALMLSDVWGSGIVIRIGICGFPVNPITDELSSSDLN	223	
QY	199	LVAAGPKSQIVMLASAEENITQDPFGAIV----	GVTQTOIIOGLOLVKETGVYKR	254
DB	224	LIVA-CTRDKTMMDVQSRSEKDLAALRLAPRAVKYIDPQI----	RLABKAGQKK	278
QY	255	TPQKLTFPSPFIVYTKTKLAMERLYAVFTYEHDKVSRDEAVNKIRIDTEQLKEKPEEA	314	
DB	279	-EYLSMLSDKTEKVLADLATRISVFTDPSYKFGKRGELDNIIGDVRVVEEEDQ	337	
QY	315	DVEIISFNVVAKEVRSIVLNEYKCDGRDLTSLRNVSGEYVMFTLGSALFQGGT	374	
DB	338	SLSLPKAVDVRKKVRSKMSISDFVDSGHVDEVPFYCESHYLPALMSALFSKDT	397	
QY	375	QVLCTVFDSLESGIKSDOVITAIINGIKDKNFMILHYFPPYATNIEIGKVTGLNRELGHG	434	

Db 398 QVLTCTVTLGAPAEAKSLSLSLV---GPPKRFMLHYSEPPPYCTNEVCGKGLNRENGHG 453  
Qy 435 ALAEKALYVPVLRD--PFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGV 492  
Db 454 TLAKALLAVLPPEAPPYTIRINSEVWSSDGTSMSVCGSMAIMADAGIPLAHVAGV 513  
Qy 493 AIGVLTDPKPKGEIEDYRLTLTDLIGEDVNGMDPKIAGTKNGKITALQADIKLPGPIK 552  
Db 514 SVGLITVDPSGSEIKDYRIYTDILIGEDHIGDMDFKAGTRDGVTAQLDRIKPAIGPLD 573  
Qy 553 IYMEAIQOASVAKKEIILQIMNKITSKPRASKENGPNVETQVPLSKRAKVFVGYNLK 612  
Db 574 IVCSLEKARAKRIQILDMKERNINSPRGDGAVSPLATIKYSNDSILRTIGPWGILKR 633  
Qy 613 KLAQETGVTLSQVDEETFPSPAPTPSVHNEAR---DFTTEICKDDQEOQLFEGAVYAT 669  
Db 634 KIEVETGARLS--LNGTLTIVAQNQDWEKAQEOVDIFI-----GRELVVGVGYYGTV 685  
Qy 670 TEIRDGVVMKLYRNMTAVLLHNTQLDNE 698  
Db 686 SSIXEYGAFAE--FPGQOGLHMSLSHE 713

## RESULT 2

polyribonucleotide nucleotidyltransferase [imported] - Clostridium acetobutylicum  
C/Species: Clostridium acetobutylicum  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: B97123  
R/Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: B97123  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-703 <KUR>  
A/Cross-references: UNIPROT:Q97145; GB:AE001437; PIDN:AAK79773.1; PID:g15024781; GSPDB:C  
A/Experimental source: Clostridium acetobutylicum ATCC824  
C/Genetics:  
A/Gene: CAC1808  
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 33.9%; Score 1204.5; DB 2; Length 703;  
Best Local Similarity 39.6%; Pred. No. 3.4e-67;  
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;  
Qy 33 DLGRKLEISSGKLARPADGSAAVQSGDTAVMTAVSKTKPSG-QFMPLVVDYRQKAA 91  
Db 8 DIAGRKLKVECGKTGM.LSNCAMFISYGDVVMMVNVNASEKREBGIDFPPLSIEYERQYS 67  
Qy 92 AGRIPTYNLRREVTSDEKELTSRLIDRSIRLPAGFYDTQVLCNLAAVGNPDLV 151  
Db 68 VGKIPGCFVKEGPESEKSIHARALDRPLRLPFGKTRNDVQVCTVWSEQNLPIL 127  
Qy 152 AINGASVALSLSDIPMNGPVAVRIGIIDGEYVNPTRKEMSSSTLNLVAGAKSQIVM 211  
Db 128 AMNGASVALSLSDIPFTTPVATVSGCIDGKFLVNPFLTEBEKSSLDLTVG-AINERYMM 186  
Qy 212 LEASAEINLQODFCHAIKVGYTQOIIGIQQVLKETGVTKRPQKLPFSPPIVYKTH 271  
Db 187 LEAQADEIPEDLMTAIDFGFNACODIYAFQEKAMKEGKEKVTPE-LYHKEEIEKQVT 245  
Qy 272 KLAMERLYAVFTYDHEKVSDEAVNKRILDTESQLEKPEADPYEIISEFNVAKEVF 331  
Db 246 EFAEESIKELM-YITPDERNLRLREIKESINSEPAKPYD-DGADIDEVAVYTLQKKV 302  
Qy 332 RSIATNEKRCDCGDLTSLRNVSCEVDMFKTLHGSALFORQOTVLCVTFPDSLESGIKS 391  
Db 303 RNMILKHEHRRDGRFPDEIRPISCDVLLPRTHSGSLFTRGLTQVMTYTL-----GPIGD 358  
Qy 392 DQVITAINGIKDKNFMILHYEPFPYATNIEIGKVTGLNREELGHGALAEKALYVPIR-RDF 449

Db 359 AQVTDIGVSESKRYMHYINFPYSTGEVRLKGPNNREIGHGALAKALVLPISSEER 418  
Qy 450 PFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVIGVTTDPEKGEIED 509  
Db 419 PYTRLVSEVLSNGSTISQASVCGSSTLALMDAGPIRRPAAIAGMIITSEDSLKEAV-- 476  
Qy 510 YRLITDLIGEDVNGMDPKIAGTKNGKITALQADIKLPGIPIKIWEAIQOASVAKKEIL 569  
Db 477 ---ITDIQLEDFEGDMDFVAGTEKGIITAIQVDTKIHGLSKYCIKTAINDARKARLFI 533  
Qy 570 QIMNKITSKPRASKENGPNVETQVPLSKRAKVFVGYNLKLAQETGVTLSQVDEET 629  
Db 534 EKWACINPEPKELSTYAPAYTINIDTDKIRTLIGGKTIKLTIEGTGKIDIREDT 593  
Qy 630 FSVAPTPSVHNEARDFITEICKDDQEOQLFEGAVYATITTEIRDGVMMKLYPNMTAVL 689  
Db 594 VFVLSDDADSANALAKMIDDLTD-----VKGEVYLGKTKTKTNFGAFVBLPGKSG-L 647  
Qy 690 LHNTQLDNERLN 701  
Db 648 WHISKLDINKVN 659

## RESULT 3

polyribonucleotide nucleotidyltransferase [imported] - Staphylococcus aureus (strain N31)  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C/Accession: B89901  
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mitutani, O.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; i  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hizamatsu, K.  
Lancet 357, 1225-1240, 2001  
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: B89901  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-698 <KUR>  
A/Cross-references: UNIPROT:Q99U08; GB:BA000018; PID:g13701074; PIDN:BAB42369.1; GSPDB:G  
A/Experimental source: strain N315  
C/Genetics:  
A/Gene: pupA  
C/Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 33.5%; Score 1190.5; DB 2; Length 698;  
Best Local Similarity 39.1%; Pred. No. 2.5e-66;  
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;  
Qy 24 SAGRAVAVDLAGNRKLEISSGKLARPADGSAAVQSGDTAVMTAVSKTKPSGQFMPLV 83  
Db 2 SOEKVKTETMAGRSLTIERGQAKONGAVLVRYGDTVLLSTATASKEPRDGFPLTV 61  
Qy 84 DYRQKAAAARIPNYNLRREVGSDEKELTSRLIIDSIRLFPAGYTPYDQVLCNLAAV 143  
Db 62 NYEKMYAAAKIPGCFKRGKRGDDATLTARLIDREIRLPFGKYGKNDVQIMMVLSD 121  
Qy 144 GVNEDVLAINGASVALSLSDIPMNGPVAVRIGIIDGEYVNPTRKEMSSSTLNLVAG 203  
Db 122 PDCSPQMAAIGSSMALSVSDIPQGPDIAGVNVGIDGKIINFTYBEKVSRLDLEAG 181  
Qy 204 APRSQIVMEASAEINLQODFCHAIKV-----VYKTOOIIGIQQVLKETGVTKRP 256  
Db 182 -HDAVAVMVAGASEITEQEMLEAIFGHEIGRLVDFQOQYDHIQPVKOE----- 232  
Qy 257 QKLPFS-----PEIVKTKHLAMERLYAVFTYDHEKVSDEAVNKRILDTESQLEK 309  
Db 233 ---FIPEARDEALVERKYSLTEEKGKLETVLT-----DKQORDEMLDNK---BEIVNE 281  
Qy 310 KPEPADP-----YEIESNVVAKVEFRSIVLNEYKRCGRDLTSLRNVSCEVDMFKTL 363  
Db 282 FIDBEDENELIKVEYAILNELVKEVRLLADEKIRPDGRKRPDEIRLPDSEVGILPRT 341

```

Qy 364 HGSAALPFGQTQVCTVTFDSLESIGKSDQVITAINIGIKDNFMHYEPFPATNIEIKV 423
Db 342 HGSGLFRGTQVCTVTFDSLESIGKSDQVITAINIGIKDNFMHYEPFPATNIEIKV 397
Qy 424 TGLNRELIGALAEKALYVIP--RDPFTIRVTSVLESNGSSSMASACGSLALMDS 481
Db 398 RAPRRRIGHALGERALKYIIPDADPPYIRIYSEVLESNGSSSMASACGSLALMDS 457
Qy 482 GVPISAVAGVAGLVNTPDEKGBIDYRLITDILGIEDNMGDMDFIAGNNGITATQ 541
Db 458 GVPKAPVAGVAGLVNTPDEKGBIDYRLITDILGIEDNMGDMDFIAGNNGITATQ 510
Qy 542 ADILPGPIKIVMEALIQASVAKKEILQIMNKITSKRPAKENGPPVETVQVPLSKRA 601
Db 511 MDIIDLGTREIRIEALEQARGRLIENHMLQITDQRTLSAAYAPKVTMTIKPKIR 570
Qy 602 KFPVPGGVNKKLOAETGVTSIQVDETFVAPLPVMEHARDPITEICKDQEQULEF 661
Db 571 DVIPGCKKINRIIDETGVKLDIEDQGTIFIGAVDQAMINRAREIIEITR-----BAEV 625
Qy 662 GAVVTATTEIRDTGVWVKYIPNMTAVLLHNTQDNEBL 700
Db 626 GQYQATVTKRIEKYGAFLPGKDA-LHLSQISKNRI 663

```

## RESULT 4

```

AE2586
polyribonucleotide nucleotidyltransferase pnpA [imported] - Agrobacterium tumefaciens (A
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AE2586
R/Mod: D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.
erge, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutayarin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Komet, F.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreepan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577, MUID:21608550, PMID:11743353
A/Accession: AE2586
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-713 <KUR>
A/Cross-references: UNIPROT:Q8UJ56, GB:AB008688, PDB:1AAL1107.1, PDB:1G1738399, GSPDB:C
A/Experimental source: Strain C58 (Dupont)
C/Genetics:
A:Gene: pnpA
A:Map position: circular chromosome
C/Superfamily: polyribonucleotide nucleotidyltransferase

```

```

Query Match 32.8%; Score 1166.5; DB 2; Length 713;
Best Local Similarity 39.1%; Pred. No. 8.2e-65;
Matches 265; Conservative 133; Mismatches 251; Indels 29; Gaps 12;

Qy 29 AVALDLGRKLEISGKLARFADGSAVVGSDTAVMTAVASKTSPSQ--FMPLVVDYRQ 87
Db 6 SVEIEMAGRPKLTGTGVARQADAVATYGETVNLATVSAKSPKQDPPFLTVNQE 65
Qy 88 KAAAGRIPTVYLRREVTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGYNE 147
Db 66 KTYAAGRIPGYFRREGRPSEKELTVSRILDRPRLPPEGYKNDTVQVVVVIQHDLEND 125
Qy 148 PDVLAINGASVALSLDIPMNGPYGAVRIGIIDGYVYNPFRKEMSSSTLNLVYAGAPKS 207
Db 126 PDVLSMVAASALTLGIPFMGPGVAGVYINGEYVNLPHLDEMDSEVLDLVVAGT--QD 184
Qy 208 QIVWLEASANIILQODFCHAIKVGKVTQOIIQGIQOLVKEGTGVTKTRPKLFTPSPEIV 267
Db 185 AVLWSEAEKLEINELMGAVMFGHGFQPIYDAI--IKLAEVAAKSPRE-FEPEDPSA 240
Qy 268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKLRLDTEBQLEKPEPA--DPYEIIIESFN 324

```

```

Db 241 LENEMGLAET-ELRTAVKITEKARYAADA VTKVKNHFLPBEAGKSPSEIGAVFK 299
Qy 325 VVAKEVPSIYLVNRYKCDGDLTSLRVNCGEVMFKTLHGSAALFORQOTVCTVTFDS 384
Db 300 HLOAKIYANVLDTKSRIDGDLSTVRIYSEVGLIPRTHSALFTKGETAIYVATIGT 359
Qy 385 LESGKSDQVITAINIGIKDNFMHYEPFPATNIEIKGTGLNRELIGHALAEKALYV 444
Db 360 GE----DEQVDSLTGMVKERFLHYNPPYPSVGETGMSGPGRREIGHGLAMRAIRFM 415
Qy 445 IP--RDPFTIRVTSVLESNGSSSMASACGSLALMDSGVPISAVAGVAGLVTKTD 502
Db 416 LPTAEPFPYTLRVVSEITSENGSSSMATVCGTSLAMDAGVPLAKPVAGIMGIL----- 471
Qy 503 EKEIEDYRLITDILGIEDVNGMDPKIAGTNKGTITLQADIKLPGPIKIVMEALIQAS 562
Db 472 ---EGERPAVLSDILGDEDLGMDPFVAGTADGTTISQMDIKIAGITEEMKALNEAQ 528
Qy 563 VAKKEILQIMNKITSKRPAKENGPPVETVQVPLSKRAKFPVPGGVNKKLOAETGVTI 622
Db 529 GGRHILGEMANALTESRGQGFAPRLVWNIIPVDKIRVIGSGKVTREIVETKAKI 588
Qy 623 SQVDETFVAPLPVMEHARDPITEICKDQEQULEFQAVYTRATITTEIRDTGVWVKY 682
Db 589 NIEDGTIVKJASASGKEIEAKRWIHSIVAEF-----EVGQIVGTVKTDAPGAFVNF 643
Qy 683 PNMTRAVLLHNTQDNEBL 700
Db 644 GARDG-LVHISQLASERV 660

```

## RESULT 5

```

P97368
polyribonucleotide nucleotidyltransferase (PA4740) [imported] - Agrobacterium tumefaciens
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: P97368
R/Goodner, B.; Hinkle, G.; Gatlung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2223-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359, MUID:21608551, PMID:11743394
A/Accession: P97368
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-713 <KUR>
A/Cross-references: UNIPROT:Q8UJ56, GB:AB007869, PDB:1AAK5903.1, PDB:1G15154950, GSPDB:C
A:Gene: AGR_C_124
C/Genetics:
A:Map position: circular chromosome
C/Superfamily: polyribonucleotide nucleotidyltransferase

```

```

Query Match 32.8%; Score 1166.5; DB 2; Length 713;
Best Local Similarity 39.1%; Pred. No. 8.2e-65;
Matches 265; Conservative 133; Mismatches 251; Indels 29; Gaps 12;

Qy 29 AVALDLGRKLEISGKLARFADGSAVVGSDTAVMTAVASKTSPSQ--FMPLVVDYRQ 87
Db 6 SVEIEMAGRPKLTGTGVARQADAVATYGETVNLATVSAKSPKQDPPFLTVNQE 65
Qy 88 KAAAGRIPTVYLRREVTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDGYNE 147
Db 66 KTYAAGRIPGYFRREGRPSEKELTVSRILDRPRLPPEGYKNDTVQVVVVIQHDLEND 125
Qy 148 PDVLAINGASVALSLDIPMNGPYGAVRIGIIDGYVYNPFRKEMSSSTLNLVYAGAPKS 207
Db 126 PDVLSMVAASALTLGIPFMGPGVAGVYINGEYVNLPHLDEMDSEVLDLVVAGT--QD 184
Qy 208 QIVWLEASANIILQODFCHAIKVGKVTQOIIQGIQOLVKEGTGVTKTRPKLFTPSPEIV 267
Db 185 AVLWSEAEKLEINELMGAVMFGHGFQPIYDAI--IKLAEVAAKSPRE-FEPEDPSA 240
Qy 268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKLRLDTEBQLEKPEPA--DPYEIIIESFN 324

```



```

Dh 241 LENEMLGAET -EIRTAVKITTEKARYAADAIVAKTKVAHFLPEEGEAKTSPEEITGA VFK 299
Qy 325 VVAKEVERSIYLINEYKRCDDGRLTSLRNVSCCEVMFTLHGSALFORGOTLOVLTCTVFD S 384
Dh 300 HLOAKIRYMNVLDRKSIDGRDLSTVRPIVSEVGLRPRTHOSALFTGEIRQAIYVATLGT 359
Qy 385 LESGICKDDOVTAINIGIKDKNFMLHYEPPIYATNEIGKVTGLNREIJGHGLAEKALYPV 444
pb 360 GE----DEQYVDLSLTGMWYKERFLLHYNPPPYSVGETGMSPGRREIGHGKLARAIRPM 415
Qy 445 IP--RDPPTIRYVSEVLESNGSSSMASACGSLALMDSGPPISSAAGVALGLVTCTDP 502
Dh 416 LPTAOGPPYTLIRVSELTETSSNGSSSMATVCTSLALMDAGVPLAKPAAGIAMGIL---- 471
Qy 503 EKEGIEYRLLTDLIGIEDYNDMDFKIAGNKGITLALOADIKLPGIPIKIYMEALIOOAS 562
Dh 472 ---EGERPAVLSDLIGDEBDLGDMDFKVAGTAGDITSLQMDIKIAGITTEIMKALIEAOQ 528
Qy 563 VAKKEIIOJMNKITSKPRASRKENGPPVETQVPLSKRAKEVGPBGYNLKKLQAEVTYI 622
Dh 529 GGRGHITIGEMANALTTESKQGLGFAPRIREVNNIIVDKIRAEVIGSGKVIIRIVETGXKI 588
Qy 623 SQVDETFESVAFPPSVNMEARDPITBEICKDOBOOLEFGAVYATITTEIRDTGVMVKLY 682
Dh 589 NIEBDGTVKIASAGKEIEARKWTHSIABP-----EVGQIYEGTVKTDALFGAFVVF 643
Qy 683 PNMTAVULLIHTQLODNERL 700
Dh 644 GARDG-LVHISQLASERV 660

```

RESULT 6  
AD0901  
polynucleotide phosphorylase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh  
C/Spectis: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A/Note: this species has also been called *Salmonella typhi*  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Oct-2003  
C/Accession: AD0901  
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
, S.; Mole, S.; O'Gaora, P.  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Oct-2003  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
A/Reference number: AB0502; MUID:21534947; PMID:11677608  
A/Accession: AD0901  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-711 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CADD7802.1; PID:916504350; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY3463  
A/Superfamily: polyribonucleotide nucleotidyltransferase

[illegible]

```

Db 191 SEAEILSBDTMLGAVVFHGEQQQVVIQAINDLYKBAKGKPRMDQ-----PEAVNDALNA 244
Qy 269 YTHKLAMRL---YAVFTDYEDHKVSRDEAVNKIRLDTBEQOLKEKPEADPYEIIIESFNV 325
Db 245 RVAAALAESRLSDAYRI-----TDKQERYAQVAVIKSETTIEQLAEBETTLANEIGETILHA 299
Qy 326 VAKEFERSIVLINEYRCDRDLTSLRNVSCEVDMFRTLGSALFORQOTVLCTVFPDNL 385
Db 300 IEKVYVRSRVLAGEPRIDOREKDMIRGLDVRGTGLPRTHGSALFTRGRETQALVATATIGT- 358
Qy 386 ESGIKSDOVITAINGIKDKNFMKLTHTPEPPYATNEIKVTGLNRELGHALAEKALYPVI 445
Db 359 ---ARDACVULBELMERTDSEFLFHYNFPYVSGETMVSQPKRREIGHGLAKRGVALVM 415
Qy 446 F--RDFFPTIRTSVELEBSGSSSMASACGSIATLMDSGVPISSAVGVALIGLTKTPDE 503
Db 416 FDMDKFPTTVKAVSIIETESNGSSSMASVCGASIALMDAGVPIKAAVAGIAMGLVKEBD-- 473
Qy 504 KGEIEDYRLTLDIIGIEDYNGDMDFKIAGTNKGITALLQADIKLPGIPIKIVMEAIQOASV 563
Db 474 -----NYVVLSDILLDEDEHLGDMDKFKVAGSRDGI$ALQMDIKIEGITKEIMOV$ALNADKG 528
Qy 564 AKKETLOMNTKISRPASRKENGPRVETVOVPL$SRBAKRVGGGNVLKTLQAE$GT$YIS 623
Db 529 ARHLHLGMEQKINAPRGDISEFAPRIRITIKISTDKIKOVIGGGSVIRLTLTE$GT$TIE 588
Qy 624 QVDEETFSVFAPT$P$VMEHARDPITEIKCHDOEQOLEFGAVYATATITIEIDTGYMVKLYP 683
Db 589 IEDGTQVIAATDGEKAKYAIRIBEI-----TATIEVGRIVNSKVTRIVDFGAFVAIGG 643
Qy 684 NMTAVLANTQOLDNERL 700
Db 644 GKEG-LVHISQIADKRV 659

```

RESULT 7  
G97782  
hypothetical protein pnp [imported] - *Rickettsia conorii* (strain Malish 7)  
C|Species: *Rickettsia conorii*  
C|Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C|Accession: G97782  
R|Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rou  
Science 293, 2093-2098, 2001  
A|Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.  
A|Reference number: A97700; MUID:21442074; PMID:11557893  
A|Accession: G97782  
A|Status: preliminary  
A|Molecule type: DNA  
A|Residues: 1-749 <KUR>  
A|Cross-references: UNIPROT:Q92HV7; GB:AE006914; PIDN:AAU03201.1; PID:G15619751; GSPDB:GR  
C|Genetics:  
A|Gene: pnp  
A|Superfamily: polyribonucleotide nucleotidyltransferase

	Query Match	32.6%;	Score 1158.5;	DB 2;	Length 749;	
	Best Local Similarity	38.0%;	Pred. No. 2.8e-64;			
	Match 257; Conservative	145;	Mismatches 244;	Indels 31;	Gaps 12;	
OY	31	AVDLGNRKLEISSGKLARFADGSAAVQSGDIAVMVTAVSKTKSPS-QFMPLVVDRÖKA	89			
Db	8	SVTMNGVLEISTGIKARIQADGAVTVKMNSVYLCTAAVVANKAKEGIGFLPTINREMA	67			
OY	90	AAAGRIPIPNYLAREVGTSPEKEILTSRIIDRSIRPLEPAGIFVDQVCINLAVGVNEPD	149			
Db	68	YAAGRIPGGFEGHEKASREVLSRLIDRPIRLPHPAVNTHIVCSVLSTDETFVD	127			
OY	150	VLAINGASVALSLSDIPMGPGVAGARIQIIDGEVYNPTKESSSTLANLVNAPAKSQI	209			
Db	128	ILATIGASAALSLSPAYLEIIVAASKVGLINGEFVNLPTLALKTSQLDVVAQTSDS-V	186			
OY	210	VMLEASANILOPDFCHAIKVGVKTTQQIIIGIQQLVKELGVKTRTPÖKLPFBPELYKY	269			
Db	187	MNVSEAHILSEBQWLEAVKFGFESFÖPIKIKELAEEKKKKLEMOÄLYPPAS-LIKE	244			

```

Oy 270 1HKL---AMELXAVFDYEDHDKXSRDAVWKIRLDTBEOLKEKPEADPBEIIESFNV 326
Db 245 IEBKFWKEIEOAPFKSKQER-STVLDLIPEKVLTTHFVSDIENK--KTSNQOISALKA 301
Oy 327 AKEXFRSIVLNEBYKRCDRDLTSLRNVSCEYDMFETLGSALFORGOVQLCTYTF-DSL 385
Db 302 ESDLIRNEILKKNRIRDRSTTDI RQINCEIIGLPSANGSALFTRGETQSLVSTFFGRSL 361
Oy 386 ESGIKSDOVITAINCIGIKDKNFMALHYEPPPYATNEIGKVTGLNRRLHGHALAEKALYVI 445
Db 362 D-----EDIVSLBEEYKERFMLNYI PFPYSVNTKEMPKASRRRVGSGKLABRAINDIL 416
Oy 446 PR--DPEPTIVTSVELESNGSSSVASACGSLALMDSGVPISSAVAGALGVTXTDPE 503
Db 417 PNKQOFPISIRVAVETTESNGSSSMATVCGSSSLMAYGVITIKAPVAGIAMGLVK----- 471
Oy 504 KGEIEDVYLLTDLIGIEDYNGDMDPKIAGTKNGKITALQADIKLPCGIPIKIYMEALIQOASV 563
Db 472 --EGKNFVYLLDILGDEDYFGDMDPKVAGTSEGITALQMDIKIGVDPKIMKVALLEQRIL 529
Oy 564 AKKELIQIMNTTISKPRASKENGVPVETQVPLSKRAKFPQPGVYNLKKIQABETGVTS 623
Db 530 GRHLIIEGMNKVYISFPNNELSKNAPSTTIIKIDKDKIRDTIIGPGKIIKEICETSGAKID 589
Oy 624 QVDETFSPVAPTSVMAHEARDPIIEICKDOBOQLLEGAVYTAITEIRDTGVVVKLYP 683
Db 590 ISDDQTVSVVASDRDKLKVALDKIKAIIVP-----EIGEIFNGTVVVKVLDGSAGFINVYG 644
Oy 684 NMTAVLHNTQOLDNERL 700
Db 645 NKDG-FVHISSEVSGERI 660

```

```

RESULT 8
H85979
polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain EF)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: H85979
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, U.; Grobeck, B.U.; Davis, N.W.; Lim, A.; Diallanca, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157.H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: H85979
A>Status: preliminary
A:Molecule type: DNA
A:Reissues: 1-734 <STD>
A:Cross-references: UNIPROT:O8X9M3; GB:AB005174; NID:q12517775; PIDN:AAG58300.1; GSPDB:G
A:Experimental source: strain O157.H7, substrain EDL933
C:GeneticB:
A:Gene: pnp
C:Superfamily: polyribonucleotide nucleotidyltransferase

Query Match          32.6%; Score 1158; DB 2; Length 734;
Best Local Similarity 40.0%; Pred. No.2.9e-64;
Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

   35 GNRLKLEISSGLARPADGSAAVQSGDTAVMTAVTSKTKPSSQ-FMPLVDYRKAAAAG 93
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    35 GQHTVTLETGMMAKQATAAVVVMSMDTAVFTTVVGKKAKKGQDFPLTVNYQERTVAAG 94
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

   94 RIRPNVYLREVEGTGDKEILTRSIITDRSTRPLPPAGYFFPDVQLCNLAAYDGNPEVDIAT 153
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    95 RIPOGFRRBRBPESGETLIARLLIDRPTRLPFPBGFAVEGVIVATVSVNVQNVPDIAM 154
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

  154 NGASVALSLSDIPMGPGAVARIGIIDGEVYNPRKEMSSSTINTLVAGAPKSQIVMLE 213
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  155 IGAASALSLSGIFPRNGPIGARVRGYINDGYVLNPTQDKSKLDPLVAGT-EAAVLMVE 213
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

  214 ASAENIIQQDFCHAIKVGVKYYTQGIIGQILVQEVETVKTPPKLTFTPSF--EIVKYT 270
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db	214	SEABLLSDQMUGAVFPHGQQQVVIQININELVKAQ----	KPRWMDQPEPVENALNARV	269
Qy	271	HKLAMERL----	YAVFTDYHDKVSDEBAVNKIRLDTBEOLKEKFEADPYEIIIESFNVA	327
Db	270	AALAEARLSDAARI-----	TDKQERYAQVADVISEITATILABEDTULDEBELGELHARIE	324
Qy	328	KEVFRSIVLNEBKRCQGRDLTSLRNVSCVDMEFKTLHGSALEFORGOTVLTCTYFDSLES	387	
Db	325	KNVARSRLVAGPRIDGRKEMDIRGLDVRTGVLPRTHGSALEFTRGEFOALVTALTLGT----	381	
Qy	388	GIKSDQVITAINGIIKDKNFMLYEPPPYATNEIGKVTGLNRLREIGHGALAEKALYVPIT-	446	
Db	382	ARDQVULDELMGERTDTFLFHYNFPYPVGEGTGMGSPKRRREIGHRLARGLAVAMPD	440	
Qy	447	RDPEFTRVMESEVLESNGSSSMASACGSLALMDGVPRISSAVAGVAILGVTITDPEKG	505	
Db	441	MDKFPYTVAVSEIYESNGSSSMASVCGASLALMDGVPRIKAAVAGLAMGLVKEGD----	496	
Qy	506	EIEDYRLTLDIGIEDYNDMDPKIAGTKIGITALQADIKLPGIPIKIVMEAIQOASVAK	565	
Db	497	---NYVVLSDIDGDDHLDMDPFKVAGSRDGIASLQMDIKIEGITEKIMQVALLNOAKGAR	553	
Qy	566	KEILOINKKITSKPRASRKENGPNVETVQVPLPSKRAKFBVPGGVNKLKLOAETVATISQV	625	
Db	554	LHILGVMEOALINAPGDISEFAPRIHTITIKINPDKIKVDIVKGGSVIATLBEETGTLTIE	613	
Qy	626	DEETPSEVAPTPSVNHEARDFTTEICKDDEOOLFEFAGVATATITEIRDTGVMVKLYENM	685	
Db	614	DDGTVKIALATDEKAKHAIIRITEI-----	TAEIENVGRVYTGKTRIVDFGAFVAGGAK	668
Qy	686	TAVLHANTQOLNBERL	700	
Db	669	EG-LVHISQIADKRV	682	

RESULT 9

E91134

polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain R10157)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C:Accession: E91134

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome reference number: A59629; MUID:21156231; PMID:11258796

A:Accession: E91134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-734 <HAY>

A:Cross-references: UNIPROT:O8X9M3, GB:BA000007, PDB:1BA37468.1, PTD:g1363518; GSPDB:G.C:Genetics:04

A:Gene: EC64045

C:superfamily: polynucleotide nucleotidyltransferase

Query Match 32.6%; Score 1158; DB 2; Length 734;

Best Local Similarity 40.0%; Pred. No. 2, 9e-64;

Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

Oy 35 GNRKLEISSGKLARPADGSAAVVGSGTAAVWVTVAVSKTKPSPQ-FMPLVVDYRQKAAAG 93

Db 35 GQHTVLTETGMARQATTAAVMTSMDDTAFTVTVVGQKKAKPGQDFPLTVNYQERTYAG 94

Oy 94 RPTNYLREVGTSPKEIITTSRIIDRSIRPLPAGYFYDTQVLGNLIAVDGNEPDVLA 153

Db 95 RIFGFFPERREGSPGETLIARLIDRPIPLPEEGGVNEQVATVAVSVAPQVNPDIAM 154

Oy 154 NGASVALSLSDIPMNGVCAVAVIGIIDGEEVNVNPTREKSSSTLNLVAGAPKSOIWMLE 213

Db 155 IGASNAALSLGIPNGPQICAAVGVYINDQVNLNPTQDELKESKLDIVVAGT-EAAVLNVE 213

Oy 214 ASAEINILQDDFCNAKGVKVTQOIIQGIQQLVKEGTGVTKRTPOKLETPSD--EIVKVT 270

```

Db 214 SEAEILSDQMLGAVFEHQOYVIONINELIVEAG---KPRMDWPEEYVNAIARV 269
QY 271 HKLAMERL--YAVFTDYEHDKVSRDEAVNKIRLDTBEOLKEKPEADPYEIIIESFNVA 327
Db 270 AALAEARLSADYRI-----TDKQERYAQVDYIKSETIATLLAABDTIDENELGETILHAE 324
QY 328 KEVARSITVLNRYCKCDGRDLTSLRVNVCEDVMFKTLHSALFOGQOVOLCTVFPDSIES 387
Db 325 KNVRSRVLAGEPRIDGREKOMIGLDVRTGVLPRTHSALFTGTGETALVTALVTLTGT--- 381
QY 388 GIKSDOVITTAINGIKDKNFMPLYEPPPYATNBEIKVTLNRRRELGHGLAAEKALYPVP 446
Db 382 -ARBAQVDELMEGERTDTFFLPHNPPYSVBETGNVSPKRRREIGHGLARGLAVAMPD 440
QY 447 -RDFPPIRTVTSVLNESNGSSSMASACGSLALMDSGVPISAVAGVALGLVTTDPEKG 505
Db 441 MDKPPYTVRVVSEITENSGSSSMASVCGASIALMDAGVPIKAAVAAGIAMGVKEGD---- 496
QY 506 EIEDYRLLTDLIGEDYNNQMDPKIAGNKGITLQADIKLPJPIKIVMEALIOQASVAK 565
Db 497 --NVVLSTDLIGEDHLGDMDPKVAGSRDISALQMDIKETIGTKELIMOVALQAKAR 553
QY 566 KEIIOINKTISKPRASKENGPNVETOVPLSKRAKRVGCGYNLKKLQAAETGTTISOV 625
Db 554 LHIGVMEQALNARGDISEFAPRIHTIKIMPDKIKOYIGGGSVIRALTBETGTTIIE 613
QY 626 DEETFSVFAPTPSVMEHARDPITBEICKDOBOOLEFGAVYATTTIEIRDTGVMVKLYPNM 685
Db 614 DDGTVKIJAATGEKAKHAIRRIEEI-----TAEIEVGVRVYTKVRIVDGAFVAGIGGK 668
QY 686 TAVLLHNTQDUNERL 700
Db 669 EG-LVHISQIADKRV 682

```

```

RESULT 10
G75320
polynucleotide phosphorylase - Deinococcus radiodurans (strain R1)
C:/Species: Deinococcus radiodurans
C:/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
R:/Accession: G75320
S.: White, O.; Eksen, J.A., Heldelberg, J.F., Hickey, E.K., Peterson, J.D.; Dodson, R.J.; Mq.
M., Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mcg.
S.: Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:/Reference number: A75250; MUID:20036896; PMID:10567266
A:/Accession: G75320
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-810 <WHL>
A:/Cross-references: UNIPROT:Q9RSR1; GB:AEO02042; GB:AEO00513; NID:g6459848; PIDN:AAF1160
A:/Experimental source: strain R1
C:/Genetics:
A:/Gene: DR2063
A:/Map position: 1
C:/Superfamily: polynucleotide nucleotidyltransferase

Query Match      32.5%   Score 1156; DB 2; Length 810;
Best Local Similarity 38.1%; Pred. No. 4.5e-64;
Matches 261; Conservative 138; Mismatches 236; Indels 50; Gaps 14;

Cy 34 LGNRKLEISSCKLARFADGSAAVOSGDPAWVTAVASKTSPSQFMPLVVDYRKAAAAG 93
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 41 LGRNELSTETGKLAKLVSGSVTVRGDTLLLTVAQASDTQSKDPLPLTFEERHYAVG 100
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Oy 94 RIPNNYLREAVGTSDXKELLTSRIIDRSIRLPFPACGYFYDYVLGNLNAVGVNEBDVIAT 153
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 101 KIPSPFORRRERPGEKALISARITDRQIRPFPFKYGRHETOVIIITVASDAQNPADVILGP 160
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Oy 154 NGASVALSLSDIIPNMVGCAVRIGITIDGEYYVNPFRKMSSSTNLTVVAGA PKGIYWLE 213
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

```

Db      161  IGAANAALISDISIPMAQPLTCAVCVQIDGQYVNPPTTEQLTRSHMDLVAGT-REAVNMVE 219
Oy      214  ASAENILQODFCFHAIKVGVKYTOQIITQGIQOLVKEFG-----VTKRTPOKLPFPSEIY 267
Db      220  CGAQTVSSEDDLVGAIEFAHAEMQGVYALIEOMRAEYGEHEKFNFLAEBGPANDYVP----- 274
Oy      268  KYTHLAMEBRLY-AVPTDYEHDKVSRDEAVNKR.L-----DTEBOLKEKPEAPDYELIE 321
Db      275  EUTERAKAAGLDALITTEGKKDRSARKLKALRNGLIEGYVDPPTAEGSAELTQA---LKD 330
Oy      322  SFNVAKAEVFRSIVLNEYKRCDCGRD.LTSLRNVSCEVDMFKTLIGSALFORGQITVLTVT 381
Db      331  AFGKEKRELRLILIEBNRADGRGSKTVRPIIEARPLPTAGSAVFTGETGVLTGVT 390
Oy      382  PDSLESIGKSDOV---ITAINGIKDKFMLEYEPFPYATNEIGKVTGLARREIGHGALA 437
Db      391  L-----GTERDEIILLDDLTAESGDK--FLHNFPPSYSGEYKMGOGOSRREIGHGULA 442
Oy      438  EKALYLPVP-RDPFTIRVTSEVLESNGSSSMASACGSLAMDGSPVLSAAYAGVALG 495
Db      443  KRAIRAVLPSEFEFPYVIRVAGDVLESNGSSSGVTCAGTSLMDACVPLKAPAGVAMG 502
Oy      496  LVTTKDPEKEJEDYRLDLIGIEDYNGMDPKIAGTNGKIGTALQDIDKLPGPIKIM 555
Db      503  LVMEED-----NTRVLTDIIGLEBALQDMDFKVGCTAGVATLQMDIKVGGITTPQIMR 555
Oy      556  EAIQOASVAKKEILQIMNKTISKPRASKRENGFVETVOVPLSKRAKFAVGPQYNLKKLQ 615
Db      556  EALAQAQGRHLHILCKMAEVLAPRAELSPTAPHILSLKINPELIGKVIQGGQVRELE 615
Oy      616  AETGVITICQVDEEFPSPAPTPSVHMEARDFTTEICKDQEQULEFQAVVTAITTEIRDT 675
Db      616  A-MGAQVITIEEDGTVIRIFSGASGESAEAVKARIEAVTK-----EAKVGEFEFGIVAKIAPF 669
Oy      676  GVMVULYXPMATVALHNTQOLDNERL 700
Db      670  GAFVNLFPQDQD-MLHISQISEQRV 693

```

RESULT 11

S74509  
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Synecchocystis sp. (strain 7806)

N/Alternate names: protein slll1043

C/Species: Synecchocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004

C/Accession: S74509

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Oikawa, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis sp.

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S74509

A>Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-718 <KAN>

A/Cross-references: UNIPROT:P72659; EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BA416663

A/Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Genetics:

A/Gene: pnp

C/Superfamily: polyribonucleotide nucleotidyltransferase

C/Keywords: nucleotidyltransferase

Query Match 32.4%; Score 1153; DB 2; Length 718;  
Best Local Similarity 37.5%; Pred. No. 5,8e-64;  
Matches 257; Conservative 131; Mismatches 261; Indels 36; Gaps 10;

QY 31 AVDLGNRLKLEISSGLARFADGSAVVGSDTAVMTATSVSKTKSPSQFMPLVDYRQAA 90  
::|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB 7 SISDFGRDIRLMKGTLPAGSGVLIGSDTAIVLATRAKRGDGIPLFLTVDYEGRLX 66  
QY 91 AAGRPTMYLRREVSTSKELLTSRIIDSIRLFFAGYIYDPQVCLNLAVGVNEBDV 150

```

Db      67 AAGRPGGFLRREGGPPREKATLISRLIDRLPLRPLRPHMLRDBLQVATTLMSDEEVPDV 126
Qy      151 LAINGASVALSLSDIPMNGPVGARIGIIDEYVNPFRKEMSSSTLNLVAGAPKSOIV 210
Db      127 LAVGASVAVILAOI PFKGPMAAVAVGVGDPIINPYREVNNDLILVAGTP-AGIV 185
Qy      211 MLEASAEIILQODFCHAIKGVKTYQOIIQGLQVKEGTGKTPQKLFPPSP---EIV 267
Db      186 MVEGAGNLPQODIIEALDFGEAVQDILINARELMTLGLTILATSE----PPVNTAVE 241
Qy      268 KYTHKAMERLVAFTDYEHDKVSRDEAVNKIRLDTESQLKEKPEADPYE----- 318
Db      242 EFINARSKKIITVLGFDLGKGDGDAIDELIKATEVETALAEIPETOPVKSVEEDPKL 301
Qy      319 IIESFNVAKEVFSIVNEKRCGROLTSLRNVSCEVDMP-KTLHGSALFORQTOVL 377
Db      302 VGNLYKALTKKLMRKQIYDDGVRDGRKLEQVRPISCEVGLPRRVHSGGLFNRGLTQVL 361
Qy      378 CTVPFDSLSEGIKSDQVITAINGIKDKNFMHYEPFPATNIEIGKVTGLNRELGHALA 437
Db      362 SLAT---LGSFGDAODLDDLHPEDEKRYLHHYFPYISVGEARPMKSPGRREIGHALA 418
Qy      438 EKALYVPVP--RDPFFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAG 495
Db      419 ERATIPVLPQGEDPVPYVAVSVLENSGTSMSGSLTALMDAGVPIKKPVGALMG 478
Qy      496 LVTKDEPKGIEDYRLTDLIGIEDVNDMDFKAGTNKIGITALQADIKLPGIPIKIVM 555
Db      479 LIXGSD-----BIRILITDIOGIEDFLGDMDFKAGTDSGITLALQDMKIDGLSMEVVS 531
Qy      556 EAIQOASVAKKEIIOIMNKTISKPRASKENGVPVETQVPLSKRAKFGVGGVNLKTLQ 615
Db      532 KALQALPARKHILDKMLATTIRBPPELSPPAPRILITIKIEBHIGAVIGEGKTIKIT 591
Qy      616 AETGVTSIQVDETFVSFAPTPSVHMEARDFTEICDQOEQLEFGAVYATATIEIRDT 675
Db      592 EQTSCKIDIDADGVTITASSEGEARERARQMIYNTNR-----KLNESGVYIGVTRILPI 646
Qy      676 GVMVLYPNMTAVLLANTOLDNERL 700
Db      647 GAFVEVLPGKEG-MIHISQLTEGRV 670

RESULT 12
H65106 polynucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Escherichia coli (E
N:Alternate names: polynucleotide phosphorylase
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence revision 23-Jan-1998 #text_change 09-Jul-2004
C:Accession: H65106; A26118; E23984
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65106
A:Stature: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-734 <BLAT>
A:Cross-references: UNIPROT:P05055; GB:A000397; GB:U00096; NID:G2367199; PIDN:AC76198.
R:Experimental source: strain K-12, substrain MG1655
R:Regnier, P.; Grunberg-Manago, M.; Potlter, C.
J. Biol. Chem. 262, 63-68, 1987
A:Title: Nucleotide sequence of the pnp gene of Escherichia coli encoding polynucleotide
in S1.
A:Reference number: A26118; MUID:87083499; PMID:2432069
A:Accession: A26118
A:Molecule type: DNA
A:Residues: 'M', 25-379; 'R', 381-472; 'S', 474-734 <REG>
A:Cross-references: EMBL:J02638; NID:G147744; PIDN:AAA83905.1; PID:G551833
R:Evans, S.; Dennis, P.P.
Gene 40, 15-22, 1985

```

```

A:Title: Promoter activity and transcript mapping in the regulatory region for genes enc
A:Reference number: A23984; MUID:86137413; PMID:3005122
A:Accession: E23984
A:Stature: preliminary
A:Molecule type: DNA
A:Residues: 'M', 25-85 <EVA>
A:Cross-references: GB:M14455; NID:G147747; PIDN:AAA24596.1; PID:G147749
A:Note: the authors translated the codon GCC for residue 74 as Arg
C:Genetics:
A:Gene: pnp
A:Map position: 69 min
A:Start codon: TTG
C:Superfamily: polynucleotide nucleotidyltransferase
C:Keywords: nucleotidyltransferase
F:25-734/Product: polynucleotide nucleotidyltransferase alpha chain #stature predicte

Query Match      32.4%; Score 1152; DB 2; Length 734;
Beat local similarity 39.9%; Pred. No. 6.9e-64;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;

Qy      35 GNRKLEISSGGLARPADOSAVVQSGDPTAVVMTAVSKTPSPSO-FMPLVVDYRQKAAAG 93
Db      35 GQHTVLTETGMMAQATVAVNVSMDDTAVFVTVGQKAKGQDFPLTVYQERTVAG 94
Qy      94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFDTQVLGNLAVDGVNEPDVLA 153
Db      95 RIQSPFRREGRBEGETLIRLIDRPIRPLFPFGFVNEVIVATVSVNVQNVADIVAM 154
Qy      154 NGASVALSLSDIPMNGPVGARIGIIDEYVNPFRKEMSSSTLNLVAGAPKSOIVMLE 213
Db      155 IGASALALSLSGIPNPGPIGARVGVINDQVYLANPQOELKESKIDLVAAGT-EAAVLMVE 213
Qy      214 ASAENILQODFCHAIKGVKTYQOIIQGLQVKEGTGKTPQKLFPPSP---EIVKYT 270
Db      214 SEAOILSEDMQLVAGVSGHEQOQVQIOMINELVAG---KPRMDQPEEVNENLNRV 269
Qy      271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTESQLKEKPEADPYEIIESFNVA 327
Db      270 AALAEARLSDAYRL-----TKQERYAADVYKSETTATLAEDETLENELGETLHAIE 324
Qy      328 KEVERSIYLVNEYKCDGDLTSLRNVSCEVDMFTLHGSALFORQTOVLCTVPFDSL 387
Db      325 KNVVRSRYLAGEPRIDGSEKDMIRGLDVRTGVLPRTGSALEFTRGEOALVTATLGT--- 381
Qy      388 GIKSDQVITAINGIKDKNFMHYEPFPATNIEIGKVTGLNRELGHALAKALYVPVP- 446
Db      382 -ARDAQVLDLIMGRRTDPLFPHYFPYISVGETGVGSPKREIIGHGLARGVLAAMPD 440
Qy      447 -RDPFFTIRVTSVLENSGSSSMASACGSLALMDSGVPISSAVAGVAGVLTCTDPEKG 505
Db      441 MDKEPYIVRVVSKITTESGSSSMASVCGASLALMDAGVPIKAAVAGIANGLVKEGD--- 496
Qy      506 EIEDYRLTDLIGIEDVNDMDFKAGTNKIGITALQADIKLPGIPIKIVMAIQOASVAK 565
Db      497 ---NYVVLSDILGDEHLDGMDFKVAGSRDQISALQMDIKIEGTKEIMQVALNOAKAR 553
Qy      566 KEIIQIMNKTISKPRASKENGVPVETQVPLSKRAKFGVGGVNLKTLQAEVTVISQV 625
Db      554 LIIIGVMEQALNABRGDISPPAPRIHTIKINDPIKIVIGGSGVIRALTEETGTTIE 613
Qy      626 DEETFSVAPTPSVHMEARDFTEICDQOEQLEFGAVYATATIEIRDTGVMVLYPNM 685
Db      614 DGGTVKATATGGEAKAIAIRIETI-----TAIEVGRVITGKTVIYDFAFAVAGGK 668
Qy      686 TAVLLHNTOLDNERL 700
Db      669 EG-LVHISQIADKRV 682

RESULT 13
AC3457 polynucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - Brucella melitensis
C:Species: Brucella melitensis

```

C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C/Accession: AC3497  
R/Delecocho, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Muejer, C.; Los, T.; Ivanova,  
- Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A/Reference number: AD3252; PMID:11756688  
A/Accession: AC3497  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-714 <KUR>  
A/Cross-references: UNIPROT:Q8YEB7; GB:AE008917; PIDN:AAL53142.1; PID:G17984011; GSPDB:C  
A/Experimental source: strain 16M  
C/Genetics:  
A/Gene: BMEI1961  
A/Map position: 1  
C/Superfamily: polyribonucleotide nucleotidyltransferase  
C/Keywords: nucleotidyltransferase

Query Match 32.3%; Score 1150; DB 2; Length 714;  
Best Local Similarity 39.0%; Pred. No. 8.8e-64;  
Matches 268; Conservative 124; Mismatches 247; Indels 48; Gaps 14;  
QY 30 VAUULGNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYRQK 88  
DB 7 VELEWGRPLTLETGKIARQADGAVLATYGETAVLATVSAKEBPFGDPFPLTVNYQEK 66  
QY AAAAGRIPTNVLREVGSDKEILTSRIIDRSIRPLPAGYFYDTQVLCNLLAVDGVNRP 148  
DB 89 AAAAGRIPTNVLREVGSDKEILTSRIIDRSIRPLPAGYFYDTQVLCNLLAVDGVNRP 148  
QY 67 TYAAGKIPGGYFKYKGRSEWETLVSRIDRIRPLPFDGKIDQVITVYVQLQDLNRP 126  
QY 149 DVLAINGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQ 208  
DB 127 DILSMVAASALTLISGVPFGPISGARVGYIDGEYVNPTRKEMSSSTLNLVAGAPKQ 185  
QY 209 IWMLEASAGENLQDFCHALIVGYKYTQOIIGQIQOLYKGTGVTYRPOKLFPTSPPELVK 268  
DB 186 VLMWSESEOLPEBVMGAVWFHKSFPQVIDAI--IKLAEVAKEBRDF--QPEDLS 239  
QY 269 YTHKLAMERLYAVPTDYEH-----DKVSRDEAVNKRIRLDTBEQKKEKPE-----AD 315  
DB 240 ---ELEAKVLA VNDLEAKVITEKQKRYAAVDAKAKAEHR--PBEVETEMSAE 293  
QY 316 PYEILIESFNVAKEVPSIVLNEYKCDGRDLSLRNVSCEVDMFKTLHGSALFQSGQTO 375  
DB 294 QFATL--FKHLQAKIVMNLIDTGNRIDGRDLSVTRPIVSEVGLPRTHSALFTRGSTQ 351  
QY 376 VLTCTVTPDSLSGSKSDQVITAINGIKDKNFMHYEPPPYATNEIGKVTGLNRRELGHGA 435  
DB 352 AIVVATLTGTGE---DEQMDLALGTYESFMLHYNPPYVSGETGRMGSPGRREIGHGX 407  
QY 436 LAEKALYVIP--RDPPTIRVTSEVLESNGSSMASCGSLALMDSGVITSAVAGVA 493  
DB 408 LAMPAIHFMPLAAQOFPTTIRAVSELTESNGSSMAVATCGSLALMDGVIIVAPVAGIA 467  
QY 494 IGLVTKTDPKEGEIEDYRLTDIIGIEDYNDMDFKIAGTKGITALQADIKLPGIPIKI 553  
DB 468 MGLIK-----EEERFAVLSDIIGDEHLDGMDFKVAGTEFGTSLQMDIKIDIGITEEI 520  
QY 554 VMEAIQOASVAKKEILQIMNTKTIKSPRAKENGVPVETVQVPLSKRAKVPFGGYNLKK 613  
DB 521 MKVLAEOAKGGRVHILGEMAKAIISSRAELGEFAPRIEVMNIPTDKIRDIVGSGGKVLRE 580  
QY 614 LQAEFTGTSIQVDEETPSVFAPTPSVMEHADFTTEICKDQEOQLERGAAYTATITIEIR 673  
DB 581 IVEFTGAKINIEDGTVKIASSNGKEIEAAKKWIHSIAEP-----EVEGTIEGVTKTA 635  
QY 674 DTGVAVKLYPMNTAVLLHNTOLDNERL 700  
DB 636 DFGAFVNFPPRDC-LVHISQLAADRV 661

RESULT 14

AC0424  
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - *Yersinia pestis* (str  
C/Species: *Yersinia pestis*  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C/Accession: AC0424  
R/Parikh, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davey, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, I  
Nature 413, 523-527, 2001  
A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: AC0424  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-705 <KUR>  
A/Cross-references: UNIPROT:Q8ZBC6; GB:AL590842; PIDN:CAC92719.1; PID:G15981414; GSPDB:G  
A/Genes: pnp  
C/Superfamily: polyribonucleotide nucleotidyltransferase  
C/Keywords: nucleotidyltransferase

Query Match 32.2%; Score 1146; DB 2; Length 705;  
Best Local Similarity 40.0%; Pred. No. 1.5e-63;  
Matches 263; Conservative 113; Mismatches 259; Indels 32; Gaps 10;  
QY 35 GNRKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQ-FMPLVVDYRQKAAAAG 93  
DB 12 GQHTVTLETGMARQATAAVVMSMDTAVFVTVGQCKAKRPGSFFPLTVNYGERTYAG 71  
QY 94 RIPNRYLRREVGSDKEILTSRIIDRSIRPLPAGYFYDTQVLCNLLAVDGVNRPDLAI 153  
DB 72 RIPSGFRREGRRSEGGTLTSRLIDRIRPLPDSFLNEQVATVAVSQVQNPQINPDVAL 131  
QY 154 NGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQIWMLE 213  
DB 132 IGASVALSLDIPWNGVAVRIGIDGEYVNPTRKEMSSSTLNLVAGAPKQIWMLE 190  
QY 214 ASAENILQDFCHALIVGYKYTQOIIGQIQOLYKGTGVTYRPOKLFPTSPPELVK 271  
DB 191 SEADILSEEQMLGAVVGHGQOVLEINALVAEAKPMDWQ-----AEPVNEALHAR 245  
QY 272 --KLAMERLYAVPTDYEHDKVSRDEAVNKRIRLDTBEQKKEKPEADPEYELIESFNVAKE 329  
DB 246 VAEIAENRILDAARITE--KQERYTQVDALKADVTETLLAQQDITLDAEIQDILASVEKN 303  
QY 330 VPSRIVLNEYKCDGRDLSLRNVSCEVDMFKTLHGSALFQSGQTOVLTCTVTPDSLSGI 389  
DB 304 VMSKRVLRGPRIDGRKMDIRGIDVARTGILPRTHSALFTRGSTQALVTATLGTAADQ 363  
QY 390 KSDQVITAINGIKDKNFMHYEPPPYATNEIGKVTGLNRRELGHGALAEKALYVIP--R 447  
DB 364 NIDELM-----GERDSEFLHYNPPYCVGETGMGSPKREIGHGRILAKKGVLAVMPSAS 419  
QY 448 DPEPTIRVTSEVLESNGSSMASCGSLALMDSGVITSAVAGVALGVTCTDPKEGEI 507  
DB 420 EFPYTIIVVSELTESNGSSMASVCGASLALMDAGVIRKAAVAGIAGLVKEG----- 473  
QY 508 EDYRLTDIIGIEDYNDMDFKIAGTKGITALQADIKLPGIPIKIYMEAIQOASVAKKE 567  
DB 474 -NFVLSLDIIGDEHLDGMDFKVAGSDGYTALQMDIKIGITREINQVNLQAKGARLH 532  
QY 568 ILQIMNTKTIKSPRAKENGVPVETVQVPLSKRAKVPFGGYNLKKLQAEFTGVTISQVDE 627  
DB 533 ILGVMEQAISTPQDIDSEFAPRIYTKMINPEKIKDVGKGSVLRALDTDEGTTEIEDD 592  
QY 628 EFTSEVPAPTPSVMEHADFTTEICKDQEOQLERGAAYTATITIEIRPTGVAVKLYPMNTA 687  
DB 593 GTIKIAATTDGAKKHAIRIEI-----TAEIEGRIVAGKYRIVDFGAFVAGGKEG 647  
QY 688 VLLHNTOLDNERL 700  
DB 648 -LVHISQLAADRV 659

Search completed: January 28, 2005, 19:35:42  
Job time : 46 secs

```
RESULT 15
G83950
polynucleotide phosphorylase pnpA [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: G83950
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83950
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-704 <STO>
A:Cross-references: UNIPROT:Q9KAB3; GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA061
A:Experimental source: strain C-125
C:Genetics:
A:Gene: pnpA
C:Superfamily: polynucleotide nucleotidyltransferase

Query Match 32.2%; Score 1145; DB 2; Length 704;
Best Local Similarity 39.3%; Pred. No. 1.8e-63;
Matches 265; Conservative 132; Mismatches 251; Indels 26; Gaps 11;

QY 31 AVDLGNRKLRTISGKLARFADGSAVVOGDTAVVWTVAVSKTPSPQFPLVVDYRQAA 90
DB 9 SIDNAGKRLTETQGLAKONGAVLVRYGDTAVLSTATASKEPKDLPFPPLTVNYEERLY 68
QY 91 AAGRIPTNYLAREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEBDV 150
DB 69 AAGRIPGFIRKREGRPESEKAILASRLIDRPIPLFPEGFRNEQVISIVMSVDQCSSE 128
QY 151 LAINGASVALSLSDIPNNGPVGAVRIGIDSEYVNPYRKEMSSSTLNLVAGAPKSGIV 210
DB 129 AAWGSSSLASISDIPPEGPVAGTVAGIDQFVINPQDLKESDIHLVVAAGT-KDAIN 187
QY 211 MLEASANILOQDFCHAIKVGVKTYQOLIIGIQQLVKETGVTKRTPQKLPFSPSEIVKYT 270
DB 188 MVEBGAEEVPEBDVNLKALMEGHNEIKRLIEQEKIAAEVGSK-TDVLVKQVDMLEQEV 246
QY 271 HKLAMERL-YAVFTDYEHDKVSRDEAVNKLRLDTEBQLEKEFPEADPYEIIIESFNVAKE 329
DB 247 RVKAEEDLKQAVQVPEKGARQDAIEAVMDKVLFTYED-NEDVPLSEVNEILHK---IVKE 302
QY 330 VPSISIVANEYKRCGCRDLTSIRNVSCVDMMKTLHGSALPQGGOTVLCVTPFDSLES 389
DB 303 EVRRLITVEKIRPDREIDRPLSSQVILPRTGSGLPFRGQTOALSTICLAL---- 358
QY 390 KSDQVITAINGIKDKNFMHYEPPEPVATNEIGKVTGLNRRRELGHGALAEKALYVIP--R 447
DB 359 GDOVQILDLGLIEESKRFMAHYNFPQFSVGETPIRGQREIRGHGALGERLLEVPIDSEQ 418
QY 448 DPEPTIRVTSEVLESNGSSSVASACGSLALMDSGVPISSAVAGVALGLVTKTDPKEGEI 507
DB 419 DPEPTIRLVSEVLESNGSTQASI-CASTLAMMDAGVPIKAVAGIAMGLV-KQD----- 471
QY 508 EDYRLRLDILIGIEDYNGDMFKIAGTKGTALQADIKLPGIPIKIYMEAIQOASVAKGE 567
DB 472 EHVSVLTDIOGMEDALDMDKFGVAGTRKGYTALQMDIKISGIDRAILLEQALEQARKGRMI 531
QY 568 ILQIMNNTISKPRARKENGSPVVEVPLSKRAKFPVGGGVNKKQDAETGVTSIQVDE 627
DB 532 ILDNMLEAISSEKSELSPPYAPKILMTINPKIRDVIGPSGKIMNKIIBDTGVKIDIBQD 591
QY 628 ETPSVFAPTPSVHAEADPITIEICKDQEQLEFGAAVYATATITEIRDTGVAVKLYPNMTA 687
DB 592 GTIYISSADTMMNKAHEIIDIVR-----EVEVGQMYLGTGVKRIEKGAFVELFKGKDG 646
QY 688 VLHNTQLDNERLN 701
DB 647 -LVHISQLAERVN 659
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using SW model

Run on: January 28, 2005, 19:28:42 ; Search time 40 Seconds  
(without alignments)

1168,855 Million cell updates/sec

Title: US-09-907-907a-42

Perfect score: 3557  
Sequence: 1 DGFPLPRDRALTOQVRA.....TAVLHNTQDNERLILP 705

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1165.5	32.8	706	3	US-09-134-001C-4908
2	1150.5	32.3	719	4	US-09-489-039A-13639
3	1140	32.0	720	4	US-09-543-681A-5738
4	1129.5	31.8	708	4	US-09-134-000C-5994
5	1126	31.7	716	4	US-09-809-665A-171
6	1107	31.1	750	4	US-09-107-532A-5868
7	1103	31.0	699	4	US-09-340-236-2259
8	1101.5	31.0	684	4	US-09-809-665A-109
9	1065	29.9	705	4	US-09-328-352-8112
10	1053	29.6	737	4	US-09-583-110-2924
11	976.5	27.5	697	4	US-09-252-991A-27283
12	884	24.9	566	4	US-09-198-452A-1073
13	860.5	24.2	332	4	US-09-270-767-44788
14	710	20.0	488	4	US-09-809-665A-44
15	636	17.9	358	4	US-09-809-665A-155
16	389.5	11.0	247	4	US-09-710-279-2006
17	340.5	9.6	91	4	US-09-270-767-60251
18	226	6.4	127	4	US-09-198-452A-1074
19	175	4.9	52	4	US-09-513-999C-6411
20	173	4.9	56	4	US-09-513-999C-7240
21	144	4.0	455	4	US-09-134-000C-6074
22	140.5	3.9	1233	4	US-09-134-000C-4971
23	140.5	3.9	1416	4	US-09-071-035-404
24	140.5	3.9	1448	4	US-09-071-035-402
25	140	3.9	252	4	US-09-252-991A-23776
26	139	3.9	921	4	US-09-800-729-199
27	138	3.9	119	4	US-09-270-767-60250

28	137	3.9	191	4	US-09-252-991A-27186	Sequence 27186, A
29	136.5	3.8	1575	4	US-09-917-254-83	Sequence 83, App1
30	136	3.8	257	4	US-09-543-681A-7481	Sequence 7481, Ap
31	132	3.7	828	4	US-09-540-236-3107	Sequence 3107, Ap
32	129	3.6	259	4	US-09-489-039A-10604	Sequence 10604, A
33	128.5	3.6	244	4	US-09-328-352-5628	Sequence 5628, Ap
34	127.5	3.6	871	4	US-09-134-001C-3979	Sequence 3979, Ap
35	124.5	3.5	471	4	US-09-107-532A-5705	Sequence 5705, Ap
36	124.5	3.5	1306	3	US-08-999-774A-13	Sequence 13, App1
37	123.5	3.5	868	4	US-09-800-729-106	Sequence 106, App
38	118.5	3.3	1781	4	US-09-961-403-13	Sequence 13, App1
39	116.5	3.3	246	4	US-09-540-236-3210	Sequence 3210, Ap
40	116	3.3	1780	1	US-08-769-309A-5	Sequence 5, App1
41	116	3.3	1780	3	US-08-994-570-5	Sequence 5, App1
42	115	3.2	875	4	US-09-107-532A-4537	Sequence 4537, Ap
43	114	3.2	648	4	US-09-198-452A-594	Sequence 594, App
44	114	3.2	2270	4	US-09-581-909-3	Sequence 3, App1
45	113	3.2	235	4	US-08-747-562-28	Sequence 28, App1

#### ALIGNMENTS

RESULT 1  
US-09-134-001C-4908  
Sequence 4908, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4908  
LENGTH: 706  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4908

Query Match: 32.8%; Score 1165.5; DB 3; Length 706;  
Best Local Similarity 38.7%; Pred. No. 2.4e-100; Indels 55; Gaps 13;  
Matches 270; Conservative 123; Mismatches 250;  
QY 24 SAGRAVAVDVGKRLKISGKLARFADGSAVVOGDTAVVTVAVSKTKPSQFMBLVV 83  
7 SGEKVKTKEMAGRSGLTERTGQLAKQANGAVLVRYGTVVLSRTAKKERDQFFLVV 66  
QY 84 DYRKAAAGRIPTNYLRREVGTSDEKILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVD 143  
67 NYEKVYAAKIRPGCFKRGKRGDEBATLRLRLDRIRPLFPAGYHDVQIMNIVLSAD 126  
QY 144 GNAEPDVLANGASVALSLDIPMNGVGVARGIIGEEVNVNPTREMSSTLANLVAG 203  
127 PDGSPFAAMAGISMAVSVDIPQGIAGVWGYDGVKVINVSADKEISRDLDELAVG 186  
QY 204 AFSQIYMLRASENILQDFCHAIKVG-----VKYTOIIOGLOQLVETGVTRTP 256  
187 -HDAVAVMVRGASSEISEMELAIPIGHIEIKLVAFQGEIIDHIOPIQE----- 237  
QY 257 QKLFTP---SPETVYKTKLAMER--LVAVFTYEHDKVSRDAVNRKRLDTBQLKEKF 311  
238 ---FVPERDEDLVEKVKSLTBKGLKDVLT---FDKQORDENLALK--EEVGHFL 288  
QY 312 PEADP-----YEISFNVAKEVPRSYLVNRYKCDGDDLSLRVSGEVDMEFKLHG 365  
289 DEEDPENETLVKEVYALINDLKEEVRLIADKIRPDGKRVKVEIRPLESEVGLPRAHG 348

QY 366 S A L F O R G G O V U C T V P B L E S G I S K S D V T A I N S I K O K N M L H Y E P P Y A T N S I G V T G 425  
 Db 349 S G L F T R G G O T A L S V L T G A L ----- G D Y O L I D B G E V E K R F M H Y N P N P S V E G T G Y R A 40404  
 QY 426 L N R E L I G H G A L E K A L Y P V I P - R D P F T I R V T S E V L E S N G S S S M A S A C G S I A L M D S G V 48303  
 Db 405 P G R R E I G H G A L G E R R A L R Y I I P T O D F P Y T I R V S E V L E S N G S S S O A S I C G S T I A L M D A G V 46404  
 QY 484 P I S A V A G V A I G V T T D P E K S I E D Y R L A T T I L E I E P Y N G D M P K I A G T K G I T A L O A D 54303  
 Db 465 P I A P A P A G I A M G L V T R D D ----- S Y T I L I D I O G M E A L D M D P K A G T A G I D G I T A L O M D 51703  
 QY 544 I K L P G I P I K T W E A I O A S V A K K E I I O I M N K T I S K P R A S R E N G S V E T V O V P L S K A F 60303  
 Db 518 I K I D G I T R E V I E R A L E O A R G O R I A I M D H L A H T I E O P R E E L S A Y A P K V T W S I N D K I R D V 57703  
 QY 604 V G F G Y N L K L O A E T V T T I S O V D E E T F S V A F P T S P M E A R D P T E I C K D O E O O L E F G A 66303  
 Db 578 I G R G G K I N I L I D E T V K I D I E Q D G I T I F I G A V D O M I R A K E I I L E D I R ----- E A V I Q 63220  
 QY 664 V Y T A I T T E I R D G V A V K Y L P Y N T A V I L A N T O U D N E R L N 70103  
 Db 633 V Y A K A V R I E K Y A F E L E P P K D A - L H I S I O S O R I N 66903

```

RESULT 2
US-09-489-039A-13639
; Sequence 13639, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Brelton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13639
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13639

```

Query Match	32.3%;	Score 1150.5;	DB 4;	Length 719;
Best Local Similarity	38.2%;	Pred. No. 6.3e-99;		
Matches 270;	Conservative 124;	Mismatches 261;	Indels 51;	Gaps 12;

```

Qy 0 LPRDRALTOLOVRALMSSAGSRAVAVDGNKKEISSSGKLARFADGSAVVOSEDVVMV 65
Db 2 IDKDRLLNMPVIRKE-----QYGHFTVLELGGMMARQTAAMVWMSMDTVFV 50

Qy 66 TAVSKTTPSPSQ-FMVLVVDYROKAALAGRIPNTYLAREVNGSDKEILTSRIIDRSIRPL 124
Db 51 TVVGQKKARCGODFFPLVYNVOERTYAACKITGCFRRBGRPSBEETLIANLIDRPPRPL 110

Qy 125 FPAGYFDYTOVLNMLIADVGVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDEYV 184
Db 111 FPEGFVNEVQVIAITVSVNPNQNPDIIVAMIGASALSLSGIPNCPIGARVGTINQYV 170

Qy 185 VNEPRKEMSSSTLNLVVAAGPKSQIVMLEASAENILQODFCHAIKVGKVTQOIIIGSIQ 244
Db 171 LNPFOELTKSSKLDLIVAGT-EAAVIVMBSBAELLESDQMIGA VVFGHEQOQVIQNIIND 229

Qy 245 LVKETGVTKRTPQCKLFTPSPELYK-----YTHKLAMEPL---YAVFTIDYHDKXSQDEAV 296
Db 230 LVKEAGKPRWDQ-----PEAVNEALNARVAALAESRLSDAYRI-----TDKQERYAQV 278

Qy 297 NKRLDPEBOLKPKPEADPEYELIESFNVAAGVRSRYIVANEYRCORDRLTSLRANSC 356
Db 279 DVKSEKSIATLVAEDETLANEGELIHAIEKNVVRSRVLGEPRIIDREKDXMTRGLDVR 318

```

```

QY 357 VDMFPTLHGSALFQCGQOVUCLTVFDPDSIESGIKSDOVTAINGIDKNFMHLYEPPPA 416
Db 339 TGVLPRTGSHLPFTKGETQALVTALVTALGTARADQNIDELM----GERTDSFLFHYNPPYS 394
QY 417 TNEIGKVTGLNRRRELGHALAEKALYPYIP-RDPFPFTRVTSVYLESSGSSMABACG 474
Db 395 VGETGMWSPKRRREIGHGRFLAKRGVLAWPTEIEPPYTVRVSELTESGSSMABVCCA 454
QY 475 SLALDGSVPISSAAVAGVAGLVTVTDBEKGEIEBYRLTDILGIEDYGDMDFKIAGTN 534
Db 455 SLALDAGVPPKAAVAGVAGLVTVTDBEKGEIEBYRLTDILGIEDYGDMDFKIAGTN 507
QY 535 KGITLQADIKLPGPIPKIVMEALIOASVAKKEIIQIMNKITSKPAASKENCPVYETQ 594
Db 508 DGISLQMDIKIEGTTKEIMOVALNOQAGARLHILIGVEQAINAPBGDISEPAPRHHTIK 567
QY 595 VPLSRKAFVGGGYNLKKLQALETGVITSQVDEETFSVPAPTPSVHMEARDPITELCKOD 654
Db 568 INPDKIKOVIGKGSVIRALTEETGTIEIEDDGTVKIATGDGKQAHARIREEI----- 623
QY 655 QEOQLFEGAVYATITTELIDTGWVWKLVPNMNAVULHNTOLMNERL 700
Db 624 -TAEIEVRITNGKVRILVDFGAFVAGIGGKEG-LVHLSQIDKRV 667

```

```

RESULT 3
US-09-543-681A-5738
; Sequence 5738; Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5738
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5738

```

Query Match	32.0%;	Score 1140;	DB 4;	Length 720;
Best Local Similarity	39.2%;	Pred. No. 6.1e-98;		
Matches 264;	Conservative 123;	Mismatches 254;	Indels 32;	Gaps 11;

```

QY 35 GNRKLEISSGKLARPADOSAVVQSDTLVMTVAASRTKSPSO-FMPLVVDYRCKAAAG 93
Db 23 GQHTLTLEGMARQATAVWMDMDGIAFVTVAAKKVKEGODFFPLTVVQSRBYAAG 82
QY 94 RLPVTYLTREVGTSDKEILTSRRIIDRSIRPLEPAGFYDYOVLCMLAADVGNPEPDLAI 153
Db 83 RLPGSFFRREGPRSGEITLIARLIDRPLRPLFPRGELNIGVATVAVSNPQVNPDIAM 142
QY 154 NGASVALSLSDIPMNGPVGAARIGIIDEYVNPFRKENSSTLNLVAGAPKSQIVMLE 213
Db 143 IGASALSLSGVPFNGPIGARGVGFIDGQVLANPITDELKISKLDLVVAGT-AGAVIAYE 201
QY 214 ASAENITLODFCHAIKVGKTKTQOIIQIGQVLKETGVTKRFPQKLFITSPELVKTH-- 271
Db 202 SEADLTSEEQMIGAVVFGHEQOOVVENINALLVAEKGKMW----WAEPE-INOGLIDR 256
QY 212 --KLAMEELVAVFTDYEDKVSRDVAANKIRLDTEBQLEKRPENDPVYIISEFNVAKE 329
Db 257 IAGLQTRTIGAVRYTE--KQERYEQIDAIRBEVISTLLAEBSLDEGEITIELFSGLEKK 314
QY 330 VFRSLVLMNEYKRCQDRDLTSLRNWSCVDWMFTLLGSALEFGOGTOVLCTVPFDSLESGI 369
Db 315 IVRAVLVAGBPPIDREKDMVWALDIRGLLPRTHGSAFTTGETOALVATILGT----A 370

```

```

QY 390 KSDQVITAINGIKDNFMHYEPFPAVATNEIGKVTGLNRELGHGALAEKALYFVID--R 447
D 371 RDAQTIDDINGEHTDTHLHNPFPYSGVETGMMGSPRRRIGGRKLAKGVLAVMPTIE 430
QY 448 DPEPTIRVTEVLESGSSSSMASACGSLALMDSGVPISSAVAGVAIGLVTKDPEKGEI 507
D 431 EFPYVAVSEITESSSSSSMASVCGASLALMDAGVPIKESVAGIAMGLVYK-----EG 483
QY 508 EDVRLTDIIGIEPYNDDMPKINGTKITLADQIKLPQIPKIYWEAIQOASVAKKE 567
D 484 ENFVLSIDILIGEDHLDMDPCKVAGSRNVSALQMDIKIEGTIEIMQVLMQKSAHL 543
QY 568 ILQIMNTISKPRASRKNQGVETVQVPLSKRAKFPVGGYNLKKLOAETGVTSQVDE 627
D 544 ILGMEASISQPRASEFAPRITIKINADKIDVIGKGSVRLATEEGTITIEEDD 603
QY 628 ETESVPAFPTSVHMBADFTIEICKDOBOOLEFGAVYTAITIEIRTVGMVKLYPMNTA 687
D 604 GTVXIATSGEQAKOALARIIEI-----TAEVGVIRYNGKVTIRIVFGAFVAIGGKEG 658
QY 688 VLHNTOLDNERL 700
D 659 -LVHISQIADKRV 670

```

## RESULT 4

```

US-09-134-000C-5994
/ Sequence 5994, Application US/09134000C
/ Patent No. 6617156
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 5994
/ LENGTH: 708
/ TYPE: PRN
/ ORGANISM: Enterococcus faecalis
US-09-134-000C-5994

```

```

Query Match 31.8%; Score 1129.5; DB 4; Length 708;
Best Local Similarity 36.6%; Pred. No. 5.8e-97;
Matches 259; Conservative 147; Mismatches 248; Indels 53; Gaps 15;

```

```

QY 9 RDRALTOLOV-RALMSAGSRAVAVDIGNRKLEISGKLAFADGSAVVOGSDTAVMVTA 67
D 1 RRRIMTEKQVFKTW-----GGRPLEVEIQQLAKQANGAVLVRYGDTVVLSAA 48
QY 68 VSKTKPSQSPQFMPFVVDYRQKAAAGRIPTMYLAREVGTSDKEILTTSRIIDRSIRPLEPA 127
D 49 VASGEAKDVDPFLLTVYEEKMTAVGKI PGCFIRKGRPSFRATILRLIDRPIRPFQ 108
QY 128 GYFYDTQVLCNLAVDQVNEPDLAINGASVALSLSDIPMNGPVAGVARIIGIIDEYVNP 187
D 109 SFNREVGITINIVMSVEQDCPEMAAMPGSSIALAISIPDPGPIAGDVGRINEYVLNP 168
QY 188 TRKEMSSSTLANLVYAKPKGOIYWLKASAKENILOODPCHAIKAVGVTKTQOIIQOIQLYK 247
D 169 TVEBAEQDIELTVAAGT-KEAINVSEGAKEVSEEDMLGALLFPDAIKELVAFQEBIYA 227
QY 248 ETGVTQKTPQCLFPTSPBIYVYTKLAMERLY-----AVFTDYHDKVSRDEAVNKIRLD 302
D 228 AVGPAPDNDVLL-----QVDAOLKKEIFDAYNTMTKTAVMTE---EKLAAREVEIDKQDT 279
QY 303 TEBOLKEKPE-ADPYEIIIESFNVA-----KEVRSIVLNEYKRCDSRDLTSLNVSCEV 357
D 280 VKEVYAEKFSHEBEAQLLKVKQIADLDKDVVRREIITDKIRPDGRKLDIEIRHLSSEV 339

```

```

QY 358 DMFPTLHGALFORQOTVLTCTVPDSLESIGKSDQVITAINGIKDNFMHYEPFPA 416
D 340 SILPRVSGSLFTRGQOALSVCVL-----APLGEHQIIDL-GVQDSKRIRHHNPFQFS 394
QY 417 TNEIGKVTGLNRELGHGALAEKALYFVID--RDPPTIRVTEVLESGSSSSMASACGG 474
D 395 VGTGRAGSPGRREIGHGALGERALAOIIPSEBDFPTITRLVAEVLSESSSSSQASICAG 454
QY 475 SLALMDSGVPISSAVAGVAIGLVTKDPEKGEIEDRLTDIIGIEPYNDDMPKINGTN 534
D 455 TLALMDAGVPIKAPVAGIAMGLVS-----DDENTYITLDIGLGBDHLDMDFKVAGTK 507
QY 535 KGTALOADIKLPQIPKIYWEAIQOASVAKKEILQIMNTISKPRASRKNQGVETVQ 594
D 508 DGIITLQMDIKIQITQITQILTEALDQAKKAMEILEETITIAAPRELSQYAPKIMIQ 567
QY 595 VPLSKRAKFPVGGYNLKKLOAETGVTSQVDEETESVPAFPTSVHMBADFTIEICKD 654
D 568 IKPAKIDVIGKGETINSIIDETGVKIDIDQGNVSIASSDAMIKALKIIEELTK-- 625
QY 655 QBOOLEFGAVYTAITIEIRTVGMVKLYPMNTAVLHNTOLDNERLN 701
D 626 ---EVEVGQVYLAQVRIEKGAFVNLIKGKDQ-LIHISQIANERVN 668

```

## RESULT 5

```

US-09-809-665A-171
/ Sequence 171, Application US/09809665A
/ Patent No. 6790950
/ GENERAL INFORMATION:
/ APPLICANT: Lowery B., David, et al.
/ TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
/ TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
/ FILE REFERENCE: 28341/00435
/ CURRENT APPLICATION NUMBER: US/09/809,665A
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: 60/153,453
/ PRIOR FILING DATE: 1999-09-10
/ PRIOR APPLICATION NUMBER: 60/128,689
/ PRIOR FILING DATE: 1999-04-09
/ PRIOR APPLICATION NUMBER: 09/545,199
/ NUMBER OF SEQ ID NOS: 197
/ SOFTWARE: Patent ver. 2.0
/ SEQ ID NO 171
/ LENGTH: 716
/ TYPE: PRN
/ ORGANISM: Pasteurella (Mannheimia) haemolytica
US-09-809-665A-171

```

```

Query Match 31.7%; Score 1126; DB 4; Length 716;
Best Local Similarity 37.1%; Pred. No. 1.3e-96;
Matches 251; Conservative 131; Mismatches 255; Indels 40; Gaps 12;

```

```

QY 35 GNRKLEISGKLAFADGSAVVOGSDTAVMVTAVSKTKPSQ-FMPLVVDYRQKAAAG 93
D 11 GQHTVLTETARARQARAAVMAAMMDTIVTVYVAKQDVAGCGDFPLTYDYQERTYAG 70
QY 94 RIPTNYLAREVGTSDKEILTTSRIIDRSIRPLFPAGYFYDTQVLCNLAVDQVNEPDLAI 153
D 71 RIFGPFKRGSRSEGETILRLIDRVRPLPFGFENELQVATVAVSNPQISPDIVAM 130
QY 154 NGASVALSLSDIPMNGPVAGVARIIGIIDEYVNPTRKEMSSSTLANLVYAKPKSOIYMLE 213
D 131 IGASVALSLSGVFNGEIGARVGFINDQFVLPNTYSEQKISRDLVVSSTDRA-VLMVE 189
QY 214 ASAKENILOODPCHAIKAVGVTKOIIQOIQLYKETGVTK--RTPOKLTSPBIYKYT 270
D 190 SEADILTEGMLAAVAVPGHQOQVVIENIKPEPKKAKPMDVABE-----PMTDLINKV 245
QY 271 HKLAMERLYAVFTDYHDKVSRDEAVNKIRLDTEBOLKEKFPADPYEIIIESFNVAKEV 330
D 246 KALAEITRLGDAYRIVE--KQVRYEQIDAIKAVYLAQLTAEDTVSEGTIIDITTALESQI 303

```

QY 331 FRSLVNEKRCGRDLTSLRNVSCEVDMFKTLHGSALFORGOTVLCVTTFDSLEGIX 390  
D 304 VRSRIAGEPRIDRVDYVALDICTSVLPRTHGSALFTRGEGQALAVATL-----GTE 358  
QY 391 SD-OVTRALINCKIKKPMFLHTEPPRYATNEIGKVTGLNRRLGHALAEKALYVIP--R 447  
D 359 RDAQIDELTGEKDRFLFHNPPYSGEGRIGSPRRREIGHRLAKRVLAVMPYAE 418  
QY 448 DEPTIRVTSSEVLSSNGSSMASACGSLALMDSGVPISAVAGVAIGLVTKTPEKGEI 507  
D 419 EPPYVAVVSEITSSSSSSMASVCGSLALMDAGVPIKAAVAGIAGMLV-----ED 471  
QY 508 EDYRLITDILGIEDYNGDMDFKIACTNKGITLQADIKLPGIPIKIYWEALIQASVAKKE 567  
D 472 EKFAVLSLIDLEDHDLGMDKFAVAGTRGTVALQMDIKIGITPEIMRIALNQAKGARMH 531  
QY 568 ILQIMNKTIKPRASRKENGVPVETVQVPLSKRAKFPVPGGYNLKKLQAEIGVTISQVDE 627  
D 532 ILGWEQALPAPRADISYAPRIHTMKIDPKKIDVIGKGGATIRALTEETNTSIDIDD 591  
QY 628 ETPSVFA---PTPSVMHEARDPITEICKDQEQOLEFGAVYATATITEIRDTGVWVKLP 683  
D 592 GTVXIAATDGNAAAVMARIEIVAEV-----EVNQIYNGKVTIRVDPGAFVSLIG 642  
QY 684 NMTAVLLNHTQLDNERL 700  
D 643 KKEG-LVHISQITNERV 658

RESULT 6  
US-09-107-532a-5868  
Sequence 5868, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007

TELEFAX: (781) 893-8277

INFORMATION FOR SEQ ID NO: 5868:

SEQUENCE CHARACTERISTICS:

LENGTH: 750 amino acids

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...750  
SEQUENCE DESCRIPTION: SEQ ID NO: 5868:  
US-09-107-532a-5868

Query Match 31.1%; Score 1107; DB 4; Length 750;  
Best Local Similarity 36.8%; Pred. No. 8.4e-95;  
Matches 250; Conservative 144; Mismatches 247; Indels 38; Gaps 12;

QY 35 GNRKLEISSGKLARFADGAVVSGDTAVWVAVTSKTPSPQPMPLVYDYYKAAAGR 94  
D 54 GGRPLQIEVQOLKQANGAVLVYGGDVVLSAAVASREARDTFPPLTINYEKMTAAAGK 113  
QY 95 IPTNYLRREVSTGDKELTSTRITDRSIRPLFPAGYFDTQVNLNLAVDGNEPVDLAIN 154  
D 114 IPGGFIRGSRPSTETATLRLIDRPIRPFAGFRNEOVNTIIVNSVKTDCSPMAAAML 173  
QY 155 GASVALSLSDIPMNPGVAVRIGIIDGEVYVNPTRKEMSSSTLNLVYAGAPKSOIWMLEA 214  
D 174 GSSIALSLSDIPPDGPAGVGVNGEYVLANPTVEQAEGTDIELFVAGT-KQAINNVES 232  
QY 215 SAENILQDPCHAIKGVGVKTTQOIIGIQIQLVKETGYTKTPQKLPSPSEIVKYTHKLA 274  
D 233 GAKVSEEDMLGALLFGFDAIKELVAFQEEIVQAVG--KEKMEVTLQVDEVIK--KEI 287  
QY 275 MERLY-----AVFTVDYHDKVSDEAVNKRILDTBEQLEKEFP--EADPY---EIIESFN 324  
D 288 FDASVATMKAAVTE--EKLAREDNIEQYKIDIREYAEKFGHEDEHLLAEVQITE 344  
QY 325 VVAKEVFRSLVNEBYKRCGRDLTSLRNVSCEVDMFKTLHGSALFORGOTVLCVTTFDS 384  
D 345 DLEKDVVRELITIDKIRPDGRKLDRIPLSSSEVSLPRVHSGGLFTRGQOALSACTL-- 402  
QY 385 LESGKSDQVITPAINGIKDKNPMFLHTEPPRYATNEIGKVTGLNRRLGHALAEKALYV 444  
D 403 --APLGEHQIIDIGVEVSRFTHNFPQFSVSGSTRAGSPRRREIGHALGERALAQV 460  
QY 445 IP--RDPPFIRVTSSEVLSSNGSSMASACGSLALMDSGVPISAVAGVAIGLVTKTDP 502  
D 461 IPSEEPFPTIRLVAEVLESNGSSSQASICAGTLALMDAGVPIKAAVAGIAGMLV----- 516  
QY 503 EKEGIEDYRLITDILGIEDYNGDMDFKIACTNKGITLQADIKLPGIPIKIYWEALIQAS 562  
D 517 ---DGENYITLTDIOGLEHDLGMDKFAVAGTKGITALQMDIKIGITBEQILTEALTOAK 573  
QY 563 VAKKEIILQIMNKTIKPRASRKENGVPVETVQVPLSKRAKFPVPGGYNLKKLQAEIGVTI 622  
D 574 QARWEIILELSTIAPRELQYAPKIEMLQIEPAKIDVIGKGGDTINGIIDEVGVKI 633  
QY 623 SQVDEEFTSVFAPTPSVMHEARDPITEICKDQEQOLEFGAVYATATITEIRDTGVWVKLY 682  
D 634 DIIDGKVSTASADAEMIKAKIKIIEDLK-----EVKGEVVLGVKVRLEKGAFAVNL 688  
QY 683 PNMTAVLLNHTQLDNERLN 701  
D 689 KKGKD-LVHISQLANDRVN 706

RESULT 7  
US-09-540-236-2259

Sequence 2259, Application US/09540236

Patent No. 6673910

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI

FILE REFERENCE: 2709.2005-001

CURRENT APPLICATION NUMBER: US/09/540,236

CURRENT FILING DATE: 2000-04-04

NUMBER OF SEQ ID NOS: 3840

SEQ ID NO 2259  
LENGTH: 699  
TYPE: PR  
ORGANISM: M.catarrhalis  
US-09-540-236-2259

Query Match 31.0%; Score 1103; DB 4; Length 699;  
Best Local Similarity 38.1%; Pred. No. 1.8e-94;  
Matches 260; Conservative 119; Mismatches 254; Indels 50; Gaps 14;

QY 35 GNRKLEISGKLAFADGSAVVGSDTAVMATAVSKTPSPSQ-FMPLVVDYRQKAAAG 93  
D 16 GNCQVLETIRIAQAN-SVLVHMGGVSLVAVVVRBEALIGQFFPLTVYQKMYASG 74  
QY 94 RIFPNYARREVGSDEKELITSRITDRSTRPLFPKGYFYDPOVLGNLAVGVNRPDLAI 153  
D 75 KIPGGYKRGGRASEFTLTSRLIDRPLRPLFPGYFNEIQVTAIVTSSDKTQADIAAM 134  
QY 154 NGASVALSLSDIPNNGVAVRIGIIDEYVNPTRKEMSSSTLNLVAGAPKSOIWMLE 213  
D 135 IGASALALISPAFNGPIGARVGFINGEYVLANPTLAEKQSDLDLVVAGT-KSAVLMVE 193  
QY 214 ASANILQODFCHAIKVGVKTYQOIIOGIQOLVKEGVTKTPQKFLTPSPBIYKTHKL 273  
D 194 SEAKELSEDQMLGAVLVGHGQQQVVIDNINTPAQAVGNTK--QEFVAPA----- 240  
QY 274 AMELVAVFTDYENDKXSRBAVANKIRLDTBEOLKEKFPRA-----DPY-----EI 319  
D 241 INEELNTQLKEQPTAKVSEAYTIR-VKQDRYARLDELAAELVLADGETSDVDYADKVAOI 299  
QY 320 IESFNVAKEVFRSIVLNEYKRCGRDLTSLRNVSCEDVMPKTLHGSALFORQTOVLCT 379  
D 300 KEEFETLKTFRVNRNIIISGKPRIDGRLETRALDIOGVLPYTHGSALFTFGSTOALV 359  
QY 380 VTFDSLESIGKSDQVITAINGIKDNFMILHYEPFPYATNEIGKVTGILNRRELGHGALAEK 439  
D 360 TLTGTLT---TRDVLVDTLGTQKODHFMILHYRPHYSVGETRGEGARREIRIGRLARR 415  
QY 440 ALFVPIV-RDFPTIRVTSVLESGSSMAACGSLAMDGVPISSAVAGVAGLV 497  
D 416 GVOAMLPAAERFPPTIRVSEITSSGSSMAVCSGLSLMDGVPLKAVAGIANGLV 475  
QY 498 TKTDPKEGEIEDYLLTDILGIEDYNGMDPKIAGTNGKITALADIKLPGIPIKIYMEA 557  
D 476 K-----EGERFVLSIDILDEBHLGMDPKVAGSVNGITALQMDIKIBITADIMEQA 528  
QY 558 IQQASVAKKEILQIMNKTIKSPRASKENGVEVTVQVPLSKRAKFGVGGYNLKKLOAE 617  
D 529 LKQAHAGRHIILNMEVIATSRTEINAHAVNATITINPEKINDVIGKGAATIRQLTED 588  
QY 618 TGVITISQVDETSVFAPTSVHMEARDFTIEICKDQEOLEFGAVYTAITTEIRDTGV 677  
D 589 TGAITIDDDNGTITIFGTDKA---STRAATIQI--EATIAVEVGTVYEGTVARIVDGA 643  
QY 678 MVKLYPMNTAVLHNTOLDNERL 700  
D 644 FTIILPG-TDGLVHISQISDERV 665

RESULT 8  
US-09-809-665A-109  
Sequence 109, Application US/09809665A  
Patent No. 6790950  
GENERAL INFORMATION:  
APPLICANT: Lowery E., David, et al.  
TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
FILE REFERENCE: 28341/00435  
CURRENT FILING DATE: 2001-03-15  
CURRENT APPLICATION NUMBER: US/09/809, 665A  
PRIOR FILING DATE: 1999-09-10  
PRIOR APPLICATION NUMBER: 60/151,453  
PRIOR FILING DATE: 1999-04-09

PRIOR APPLICATION NUMBER: 09/545,199  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 197  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 109  
LENGTH: 684  
TYPE: PR  
ORGANISM: Pasteurella multocida  
US-09-809-665A-109

Query Match 31.0%; Score 1101.5; DB 4; Length 684;  
Best Local Similarity 37.8%; Pred. No. 2.3e-94;  
Matches 246; Conservative 133; Mismatches 237; Indels 35; Gaps 13;

QY 60 DTAVMATAVSKTPSPSQ-FMPLVVDYRQKAAAGRIPTYLAREVTSDEKELITSRITID 118  
D 6 DTVFVTVVAKKQVKESQDFFPLVNTQERTYAGRIPLGEPFKREGSPSGETLIALID 65  
QY 119 RSIRPLFPAGYFYDTQVLGNLAVGVNRPDLAINASVALSLSDIPNNGVAVRIGI 178  
D 66 RPIRPLFPBGYFNEIQVATVGVNPOICPDVAMIGASALSLSGVFPNGPIGARVGF 125  
QY 179 IDEGVNPTFRKEMSSSTLNLVAGAPKSOIWMLEASANILQODFCHAIKVGVKTYQOI 238  
D 126 IDQFVLNPTMNEQKQSRDLVLVAGTDKA-VLWVSEADVLTEBOMLAAVFGHQOQOV 184  
QY 239 IOGIQOLVKEGVTK---RTPQKLFPPSPBIYKTHKLAMEBLVAVFTDYENDKXSRDEA 295  
D 185 IDAKERTAGKGRMDVAP---PNTALIEVKALIAARIGEARITE--KQAYEQ 238  
QY 296 VNKIRLDTBEOLKEKFPRAEPY---EIIESFNVAKEVFRSIVLNEYKRCGRDLTSLRN 352  
D 239 IDAIKADVIAQITAEVAGEGDISBGKIVDIFTLAESQIVASRIIAGSPRIDGRTVTVRA 298  
QY 353 VSCEDVMPKTLHGSALFORQTOVLCTVTFDSLESIGKSD-QVITAINGIKDNFMILHYE 411  
D 299 LDICTGVLPHTHSGALIFTRGETQALAVATL-----GTERDAQIIDEITGERSHFFLHYN 353  
QY 412 PPPATNEIKQVTCNMBREIGHALAEKALVPIYR-RDFPTIRVTSVLESGSSMA 469  
D 354 PPPVSGETMIGSPKREIRIGHGLARGLAAVMPPLAEPYVAVVSEITSSGSSMA 413  
QY 470 SAGGSALAMDGVPISSAVAGVAGLVTKTDPKEGEIEDYRLTDLIGIEDYNGMDPK 529  
D 414 SVCGASALMDAGVPIIAAAGIANGLVK-----EDERFVLSIDILGDBHLGMDPK 466  
QY 530 IAGTNGKITALQADIKLPGIPIKIYMEALIQQASVAKKEILQIMNKTIKSPRASKENG 589  
D 467 VAGTRGTVTALQMDIKIEGITAIRIMQALNOAKSARLHIIGVMEQALPAPRADISDPAPR 526  
QY 590 VETVQVPLSKRAKFGVGGYNLKKLOAETGVITISQVDETSVFAPTSVHMEARDITE 649  
D 527 IYTMKIDPKIKIDVIGKGAATIRALTEETSTSIDIDDGVTKLAADVG---NSAKXEWAR 583  
QY 650 ICKDQEOLEFGAVYTAITTEIRDTGVNWKLYPMNTAVLHNTOLDNERL 700  
D 584 I--EDITAEVAGAVKGVKTRADFGAFVSYGNKEG-LVHISQISDERV 631

RESULT 9  
US-09-328-352-8112  
Sequence 8112, Application US/09328352  
Patent No. 6562358  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT FILING DATE: 1999-06-04  
CURRENT APPLICATION NUMBER: US/09/328,352  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 8112  
LENGTH: 705

TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-8112

Query Match 29.9%; Score 1065; DB 4; Length 705;  
Best Local Similarity 36.8%; Pred. No. 6.7e-91;  
Matches 250; Conservative 124; Mismatches 261; Indels 44; Gaps 14;

```
QY 35 GNRKLEISSGKLARFADGSAAVVGSDTAVMVTAVSKTPSPQSPMPVLDYRQKAAAGR 94
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 22 GQHVAVLETGRVAQAMTVLITMGVTVLVAVVAAPAKAQDPFLTVVYQERQVAGR 81
QY 95 IPTVYLRKRVGTSKKEILTSRIIDRSIRPLFPAGFYDTQVLCNLLAVDGVNEPVLA 154
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 82 IPGGYGRKGRSAEATLISRLIDRPIRPLFPBEGYNEIQVATVAVSSDKTMEADIA 141
QY 155 GASVALSIDIPMNGPVGAIVRIGIIDGEVYNPTREKSSSTLNLVAGAPKQIWMLEA 214
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 142 GTSALALAGTFPFGPFGAARVGLNGEYVLPNPEQASDLDLVVAGT-ESA 200
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 215 SAENILODFCHAIKVGKVTQOIIQGIQQLVKETGVTKTPQKLFPSPEIVKTHKLA 274
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 201 EAKELSEDQMGAVLFGHDEWQIAQAINEPAAAAG--KPSDWVAPA-----HN-- 248
QY 275 MERUYAVFTDYEHDKVSR--DEAVNKIR--LDT--BEOLKEKFPADPVEIISFNVA 327
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 249 -BEIRAKIKKEAFKAEKISEAVYTIKODRYALDLAHAEVAVQFVPEEDVDGIADEV 307
QY 328 KEV-FRGI---VLNBYKRCGRDLTSLRNVSCEVDMFETLHGSALFORGOTVLCVTFD 383
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 308 EDLKRIVRDNITLSGKPRIDGRDITKTVBALVQGVLEFRAGSALFTRGETALVTT 367
QY 384 SLESIGKSDQVITAINGIKDKNFMILHYEPFPYATNEIGKVTGLNRRELHGALAEKAL 443
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 368 NTRDAL----MVDPLAGTKTDNFMILHYNFPAYSGEGRSGSPRRREIGHRLARQ 423
QY 444 VIPE--DPEFTIRVTSVLENGSSSMASACGSLAMDGVPISSAVAGVAILVTKTD 501
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 424 VLPADRPFPYIRIVSDITBESNGSSSMASVCGASLSMDAGVPLKAPVAGIAMELVK-- 480
QY 502 PEKEIEIDYRLITJLIGIEDYNGMDFKIAGTNKGITALQADIKLPGIPKIWEAIOQA 561
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 481 ----EGERFAVLSLIGDEDEHLGMDRKNVANGITRLQMDIKIETGTEIMVALLNOA 536
QY 562 SVAKKEILQIMNKITISKPRASRKENGVPVETQVPLSKRAKFPVPGYNLKLQAE 621
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 537 FAGMHILINEMNKYISARPEISMAHPFEVITINPKIRNVIGKGGATIRQIETKAA 596
QY 622 ISQVDEETFSVFAPTPSVMEHARDFTIEICKDOEQLEFGAVYTTATITEIRDTGV 681
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 597 IDIDNGTVRVFGETKAAAKAIAKIQAI-----TAEVEPGKIYDGKIVRIVEFGAF 651
QY 682 YPNMTAVLHNTQLDNERL 700
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 652 MFG-TDGLHLISQISNERT 669
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 10
US-09-583-110-2924
; Sequence 2924, Application US/09583110
; Patent No. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATHO0-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
```

PRIOR FILING DATE: 1997-07-02  
NUMBER OF SEQ ID NOS: 5322  
SEQ ID NO 2924  
LENGTH: 737  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-583-110-2924

Query Match 29.6%; Score 1053; DB 4; Length 737;  
Best Local Similarity 34.4%; Pred. No. 9.7e-90;  
Matches 233; Conservative 150; Mismatches 256; Indels 38; Gaps 9;

```
QY 37 RKLKLEISGKLARFADGSAAVVGSDTAVMVTAVSKTPSPQSPMPVLDYRQKAAAGRIP 96
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 13 RELIVETGVAKAQSAAVVRVRESYTLTAAVSSKMAQTDFPFLQVNEEKYAAAGKFP 72
QY 97 TNYLRKRVGTSKKEILTSRIIDRSIRPLFPAGFYDTQVLCNLLAVDGVNEPVLA 156
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 73 GGFMRKGRPSTAITLRLIDRPIRPLFMRPAGREYQVINTVLSYDENASAPMAAFGS 132
QY 157 SVALSIDIPMNGPVGAIVRIGIIDGEVYNPTREKSSSTLNLVAGAPKQIWMLEASA 216
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 SLALSIDIPDGPPIAGVGVYDQGIINPSQEQAEQSILLETVAGT-KHAVNVESSGA 191
QY 217 ENILODFCHAIKVGKVTQOIIQGIQQLVKETGVTKTPQKLFPSPEIVKTHKLA 276
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 192 KEISEIEMELALKHAEVAVELIAFOEIVAAVGKK-----AEVELLHVDAELQAE 243
QY 277 RLVAVFDTVE----HDKVRDEAVNKKIRLDTBEOLKEKFPADPY-----EIIESFN 326
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 244 ITAAVNSDLQAVQVEKLAKEALATQVAVYBEKADHEFDRIKRVDAELIEQM 303
QY 327 AKVFSIVLNEXKRCGRDLTSLRNVSCEVDMFETLHGSALFORGOTVLCVTFDSLE 386
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 304 EHAVERRLITEDKVRPDGRKRVDEIRPLDAVVDLPVHSGSLFTFRQTOALSVLT 359
QY 387 SGKSDQVITAINGIKDKNFMILHYEPFPYATNEIGKVTGLNRRELHGALAEKAL 446
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 360 APNGETQIIDGLPEYKRRMHNPNQYSVGETGKAGARREIGHGALGEEALQVLP 419
QY 447 --DPEFTIRVTSVLENGSSSMASACGSLAMDGVPISSAVAGVAILVTKTDPEK 504
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 420 SLEFPFAIRLVAVELESNGSSQASICAGTLLMAGVPIKAPVAGIAMGLIS----- 473
QY 505 GEIEDYRLITJLIGIEDYNGMDFKIAGTNKGITALQADIKLPGIPKIWEAIOQA 564
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 474 -DGNNTVTLVDIGLEBDFHDMDFKVAAGTRDIGTALQMDIKIGITAEILTEALQA 532
QY 565 KKEILQIMNKITISKPRASRKENGVPVETQVPLSKRAKFPVPGYNLKLQAE 624
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 533 RFEILDVIENTIPEVLELAFTAPKIDTIDVDKIKIVGKGGETTIDKIAETGVKIDI 592
QY 625 VDEETFSVFAPTPSVMEHARDFTIEICKDOEQLEFGAVYTTATITEIRDTGV 684
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 593 DEGNVSISSDDAIDRAKEITAGLVREKAVDE-----VYRAKVRIEKGAFVLLF-D 646
QY 685 MTAVLHNTQLDNERLN 701
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 647 KTDALVHISEAMWTRTN 663
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 11
US-09-252-991A-27283
; Sequence 27283, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 27283  
 LENGTH: 697  
 TYPE: PR  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-27283

Query Match 27.5%; Score 976.5; DB 4; Length 697;  
 Best Local Similarity 35.6%; Pred. No. 1.4e-82;  
 Matches 228; Conservative 130; Mismatches 230; Indels 53; Gaps 15;

86 ROKAAAGRTPTNLR-----EYTSDEI-----LTSRIIRSRPLPAGFY 131  
 45 RPKATSRRCPTTRKPTPOAVSPAVSSVAVLPRRETTLRLDIPRLPPEGGMN 104  
 132 DTQVLCNLAVDGVNEPDVLAINGASVALSLSDIPMNPVGAVRIGI-IDGEVYVNPTRK 190  
 105 EVQVVCIVSTNKSDDPDIAIMIGTSALASIGIPFAGPICAARVGFHPELGIYILPTYE 164  
 191 EMSSTNLVAVAPKSOIWMLEASAEENILQODFCHAIKGVKTYQIIGIQQLVKETS 250  
 165 QLOSSSLDMVAVGT-EDAVLWVESEADELTEDQMLGAVLFAHDEFQAVIRAVKELAAEAG 223  
 251 VTK---RTPKLTFPSEIYKTHKLAMERLYAVFTVDEHKVS-----RDAVNNKIRLD 302  
 224 KPAWDMKAPAE-NVLVNAIAELGEAISQAVTI--TIKORVNRIGELRQAAVALPAGE 280  
 303 TEOQLKEFPFADYEIIESFNVAKEVFRSIVLNEYKRCGRDLSLRNVCEDMFKT 362  
 281 EE----GKFPAS---EYQDVFGLEIRYRENIYNGKRIDGRTRTRVPIRIEVLGK 333  
 363 LHGSALFORQGTQVLCVTFPDSLESIGKSDQVITAINGIDKKNFMHYEPFPVATNEIGK 422  
 334 THGSALFTRGETQALVAVATLGT---ARDAQLDLTLEGERKDAFMHYFPFVGEBCGR 389  
 423 VTGIANRELHGALAEKALYVPIR--DFPTIRVTSVLESNGSSMAACGSLALMD 480  
 390 MGSFGRREIGHGRARGVAMLPDDEFPYTIIRVSEITSSNGSSMAACGSLALMD 449  
 481 SGVPSISAVAGVAILGVTKDPKGEIEDYRLLDILGIDYNGMDPKIAGTKNGITATL 540  
 450 AGVAVKAPVAGIANGLVK-----EGEKAFLVDILDEDEHLDGMDPKVAGTKGTATL 502  
 541 QADIKLPGIPKIVMEALQOASVAKKEILQIMNTISKPRARKENGVEVTVQVPLSKR 600  
 503 QMDIKINGITEIMEIALGQALBARINILGOMNOVIAPRAELSENAPTMLQMKIDSDKI 562  
 601 AKFVPGGYNLKKLQALGTGVTISQVDEETSVFAPTSVMAHADFITEICKDOEQOLE 660  
 563 RDVIGKGATIRGICEETKASIDIEDDGSVKIYGTEKAAEAALKRLVALI-----TAAAE 617  
 661 FGAAYTATITEIRDTGVAVVKLYPMNTAVLHNTOLDNERLN 701  
 618 IGTIVGKERIVDFGAFFVNLPGKDG-LVAHSIQSDRKID 657

RESULT 12  
 US-09-198-452A-1073  
 Sequence 1073, Application US/09198452A  
 Patent No. 6559294  
 GENERAL INFORMATION:  
 APPLICANT: Glifalab, R.  
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 TITLE OF INVENTION: and treatment of infection  
 FILE REFERENCE: 9710-003-999  
 CURRENT APPLICATION NUMBER: US/09/198,452A  
 PRIOR FILING DATE: 1998-11-24  
 NUMBER OF SEQ ID NOS: 6849  
 SEQ ID NO 1073

LENGTH: 568  
 TYPE: PR  
 ORGANISM: Chlamydia pneumoniae  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: 1...568  
 OTHER INFORMATION: Xaa=unknown or other  
 US-09-198-452A-1073

Query Match 24.9%; Score 884; DB 4; Length 568;  
 Best Local Similarity 36.9%; Pred. No. 4.9e-74;  
 Matches 205; Conservative 112; Mismatches 198; Indels 40; Gaps 13;

157 SVALSIDIIPMNPVGAVRIGIIDGEVYVNPTRKEMSSSTNLVAVAPKSOIWMLEASA 216  
 8 SALAISIDIPQSNVAVAGRICIDNCGVINPTTELASSLTDVLVAGT-ENAILMIEGHC 66  
 217 ENILQODFCHAIKGVKTYQIIGIQQLVKETSQVTKRTPDKLFTPSEIYKTHKLAME 276  
 67 DFTBEQVLDALIEGHGHIYVICRQLQMOEBVGSKNLSAVVPLPA-EVLTAVKECAQD 125  
 277 RLAVFTDYDEHDKSRDAVNNKIRL-DTEQLKEFPFADYEIIESFNVA-----KE 329  
 126 KTFELFN-----IKDKVHAATFAHEIENITLKLQREDD-DLSSFNITKAACKTLKSD 177  
 330 VFRSIVLNEYKRCGRDLSLRNVCEDMFKTLHGSALFORQGTQVLCVTFPDSLES 387  
 178 TMRALINDREIRADGRSLTVRPTTETSYLPRTHGSLCTRGETQVLAICTLSEAMA- 236  
 388 GIKSDQVITAINGIDKKNFMHYEPFPVATNEIGKTVGLNRRELHGALAEKALYVPIR- 446  
 237 -----QYEDDLNGELSKFYLYQFFPPFVGEVGRIGSPRRREIGHGKLEKALSHALPD 291  
 447 -RDEPTIRVTSVLESNGSSMAACGSLALMDGCVPISSAVAGVAILGVTKDEKG 505  
 292 SATFPYTIIRIESNTTESNGSSMAACGSLALMDADAVPISSPAGIAMGILIT---DDQG 348  
 506 EIEDYRLTLTDLIGIEDYNGMDPKIAGTKNGITATLQADIKLPGIPKIVMEALQOASVAK 565  
 349 AI-----ILSDISGLEHLDGMDPKIAGSGKITAFQMDIKVEGITPAIMKKALSQAQGC 404  
 566 KEILQIMNTISKPRARKENGVEVTVQVPLSKRAKFPVPGGYNLKKLQALGTGVTISQV 625  
 405 NDILINMNBALSAVKADLSQYAPRIETMQIKPKIASVIGPGKQRIQIILEETGVQIDVN 464  
 626 DEETSVFAPTSVMAHADFITEICKDOEQOLEFQAVVTAITTEIRDTGVAVVKLYPMN 685  
 465 DLGAVISASASSAINKAKKEIIEGLV-----GEVEVGKTYGRVTSVAVAGAFVEVLPGK 519  
 686 TAVLHNTOLDNERL 700  
 520 EG-LCHISEGSRORI 533

RESULT 13  
 US-09-270-767-44788  
 Sequence 44788, Application US/09270767  
 Patent No. 6703491  
 GENERAL INFORMATION:  
 APPLICANT: Homburger et al.  
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 FILE REFERENCE: File Reference: 7326-094  
 CURRENT APPLICATION NUMBER: US/09/270,767  
 CURRENT FILING DATE: 1999-03-17  
 NUMBER OF SEQ ID NOS: 62517  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 44788  
 LENGTH: 332  
 TYPE: PR  
 ORGANISM: Drosophila melanogaster  
 FEATURE:  
 OTHER INFORMATION: Xaa means any amino acid  
 US-09-270-767-44788



Query Match 24.2%; Score 860.5; DB 4; Length 332;  
Best Local Similarity 52.4%; Pred. No. 3.1e-72;  
Matches 174; Conservative 48; Mismatches 107; Indels 3; Gaps 2;

QY 211 MLESAENILLOODCHAIKGVXKXTOQIIOGILQVLVETVTKTPQKLFPSPEIYKXT 270  
DB 1 MLEKGNVLOODILKAIKQGTREAOPTIHEIRYQKAGKQKEVEVAADDELGA 60

QY 271 HKLMERLYAVFTYEHDKVSRDAVNKIRLDTSEOLKEKPEADPYEIESFNVAKEV 330  
DB 61 RACKCKGKESRQRPRAKISRANAVNEVRNVDKMWSSPDEPESLITQFQGTERTI 120

QY 331 FRSLVLEKRCQDRDLTSLRNVSCEVDMFKTLHGSALFQRCQVLTCTVTFDSLESIGK 390  
DB 121 FRELIFERGLRCDERDYQQLENISQYDMYKPLHGSALFQRCQVCTVSLSDSQESAMK 180

QY 391 SDQV--ITAINGICDKKPMFLHYEPFPYATNIGVTLGNRELHGAALAEALPVIPRO 448  
DB 181 LDSLAALDSXXXXPSPNFMFLHYEPFPYATGSGRIQVGRREMGHGALEKSLPLTEND 240

QY 449 PPTIRVTSVLESGSSSMASACGSLALMDSGVPISSAVAGVAILGVTK--TDPKEGEI 507  
DB 241 YPFTVRLTSVLESGSSSMASVCGSLALMDAGVPVSAPAGVAILGLVTKFENDTRKL 300

QY 508 EDYRLTDLIGIEDYNGDMDFKLAGTKGITA 539  
DB 301 QDYRLTDLIGIEDYNGDMDFKAGTRKGFTA 332

RESULT 14  
US-09-809-665A-44  
; Sequence 44, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809, 665A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 488  
; TYPE: PRT  
; ORGANISM: Pasteurella multocida  
US-09-809-665A-44

Query Match 20.0%; Score 710; DB 4; Length 488;  
Best Local Similarity 36.4%; Pred. No. 8.7e-58;  
Matches 162; Conservative 87; Mismatches 170; Indels 26; Gaps 9;

QY 262 PSPPIVKTHTKLMERLYAVFTDYEHDKVSRDAVNKIRLDTSEOLKEKPEADPY---E 318  
DB 11 PNTLTIKVKAIARJGEARJRTENKHYM--NKIDAKADVIAQITRAVAEGEDISEGK 68

QY 319 IIESFNVAKEVFSIVLENYKRCQDRDLTSLRNVSCEVDMFKTLHGSALFQRCQVLTCT 378  
DB 69 IVDLFTLSEQIVASRIIAGEPRIDGRVDVVRALDICTGLPRTTHSALITTRGETQALA 128

QY 379 TTFPDSLESIGKSD-QVITAINGICDKKPMFLHYEPFPYATNIGVTLGNRELHGAALAE 437  
DB 129 VATL-----GTERDAQIITDELTERSDHFLFHYNFPYPSVGETGMISPKRREIGHRLA 183

QY 438 EKALYPIYP--RDPFTIRVTSVLESGSSSMASACGSLALMDSGVPISSAVAGVAILG 495  
DB 184 KRGVAAVMPITLAEFPYVVRVSVITTESGSSSMASVCGSLALMDAGVPIKAAVAGIANG 243

QY 496 LVTKTDPKEGEIEDYRLTDLIGIEDYNGDMDFKLAGTKGITAQALADILKPEIPIKIYM 555  
DB 244 LVK-----EDEKFWLSDILDEDEHLDGMDPKVAGTRGVTAQMDIETIEITAEIMQ 296

QY 556 EAIQOASVAKKELIQTQNMKTIISKPRASRKENGVEVTVQVPLSKRAKFGVPGYNIKLQ 615  
DB 297 IALNOAKSAPLHLTGWEOQALPAPRADISDPARITITMKIDPKKIDQVIGKGAATTBAL 356

QY 616 AETGVTSQVDEERTFSVAFPTPSVMEHARDPITEICKDQEOQLERGAVYATATITEIRDT 675  
DB 357 EETGTSIDIDDDGTVXIAAVDG--NSAKEVMARI--EDITAVEAGAVYKGVTRILADF 411

QY 676 GVMVKLYPNKTAVALNHTQIDNERL 700  
DB 412 GAFVSIQNKEG-LVHISQIAEERV 435

RESULT 15  
US-09-809-665A-155  
; Sequence 155, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809, 665A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 155  
; LENGTH: 358  
; TYPE: PRT  
; ORGANISM: Acetivobacillus pleuropneumoniae  
US-09-809-665A-155

Query Match 17.9%; Score 636; DB 4; Length 358;  
Best Local Similarity 38.3%; Pred. No. 4.7e-51;  
Matches 143; Conservative 67; Mismatches 139; Indels 24; Gaps 7;

QY 242 IOOLVKETGVTK--RTPQKLFPSPEIYKXTKLMERLYAVFTDYEHDKVSRDAVNK 298  
DB 2 IKFVKEAGKPRNDWAPF-----PNTALINQVAKALAEARIGDXYRITTE--KQARYEQIDA 55

QY 299 IRLDTSEOLKEKPEADPYEIESFNVAKEVFSIVLENYKRCQDRDLTSLRNVSCEVD 358  
DB 56 IKADVIAQLTAQOEVYSEGAIIIDITALBESIYGRHIIAGEPRIDGRVDVVRALDICTG 115

QY 359 MFKTLHGSALFQRCQVLTCTVTFDSLESIGKSD-QVITAINGICDKKPMFLHYEPFPYAT 417  
DB 116 VLPRTHSALIFTGEQALAVATL-----GTERDAQIVDELTEKSDRFLFHNFPYSV 170

QY 418 NEIGKVTGLNRRRLHGAALAEALYPIYP--RDPFTIRVTSVLESGSSSMASACGGS 475  
DB 171 GEGTGGSPKREKRIIGHRLKRGVAVLMPFAEFYVVRVSVITTESGSSSMASACGAS 250

QY 476 LALMDSGVPISSAVAGVAILGVTKTPPEKGEIEDYRLTDLIGIEDYNGDMDFKLAGTK 535  
DB 231 LALMDAGVPIKAAVAGIANGLVK-----EEKFWLSDILDEDEHLDGMDPKVAGTRE 283

QY 536 GITALQADILKPEIPIKIYMEALIQOASVAKKELIQTQNMKTIISKPRASRKENGVEVTVQV 555  
DB 284 GVTALQMDIKIEIGITPEIMQIALNOAKGARMHLSVMEQALPAPRADISDPARITITMKI 343

QY 596 PLSKRAKFGVPGG 608  
DB 344 DPKKINDVIGKG 356

Search completed: January 28, 2005, 19:36:29  
Job time : 43 secs

---

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 19:35:48 ; Search time 147 Seconds  
(without alignments)  
1732.714 Million cell updates/sec

Title: US-09-907-907A-42  
Perfect score: 3557  
Sequence: 1 DGPFLPRRDRALTOLOVRA.....TAVVLTHTQDNERLTLLP 705

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3557	100.0	705	US-09-907-907A-42 Sequence 42, Appl
2	3557	100.0	705	US-09-907-907A-44 Sequence 44, Appl
3	2285	64.2	504	US-10-408-765A-1791 Sequence 1791, Ap
4	1267	35.6	1526	US-10-437-963-117217 Sequence 117217, A
5	1204.5	33.9	703	US-10-282-122A-51965 Sequence 51965, A
6	1191.5	33.5	696	US-09-815-242-5443 Sequence 5443, Ap
7	1191.5	33.5	698	US-09-815-242-12325 Sequence 12325, A
8	1190.5	33.5	698	US-10-282-122A-44413 Sequence 44413, A
9	1165.5	32.8	701	US-10-282-122A-70948 Sequence 70948, A
10	1161.5	32.7	692	US-10-282-122A-71588 Sequence 71588, A
11	1161	32.6	711	US-10-282-122A-75529 Sequence 75529, A
12	1158	32.6	721	US-09-815-242-13768 Sequence 13768, A
13	1157	32.5	729	US-10-282-122A-61234 Sequence 61234, A

	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
	1155	1155	1155	1147	1146.5	1146	1140	1135.5	1135	1134	1134	1132	1131.5	1128.5	1128.5	1127.5	1127	1127	1126	1126	1125.5	1121	1117.5	1117	1114.5	1112	1111	1111	1109	1107	1103	1101.5
	32.5	32.5	32.5	32.2	32.2	32.2	32.0	31.9	31.9	31.9	31.9	31.8	31.8	31.7	31.7	31.7	31.7	31.7	31.7	31.7	31.6	31.5	31.4	31.4	31.3	31.3	31.2	31.2	31.2	31.1	31.0	31.0
	703	734	734	711	713	705	709	719	712	709	709	705	705	704	704	763	702	723	716	716	714	709	714	707	694	712	706	706	736	702	695	684
	US-10-282-122A-53063	US-09-815-242-10312	US-10-282-122A-56702	US-10-282-122A-59492	US-10-282-122A-77935	US-10-282-122A-77935	US-10-282-122A-68773	US-10-282-122A-51291	US-10-282-122A-56195	US-09-815-242-10989	US-10-282-122A-58109	US-09-907-907A-43	US-10-282-122A-48837	US-09-815-242-10882	US-10-282-122A-42421	US-10-282-122A-50508	US-09-815-242-4975	US-10-282-122A-60483	US-09-809-665A-171	US-10-854-299-171	US-10-282-122A-67188	US-10-282-122A-77113	US-10-282-122A-49590	US-10-282-122A-65516	US-10-282-122A-54319	US-10-282-122A-45343	US-10-282-122A-65836	US-10-275-026A-36	US-10-282-122A-57601	US-10-282-122A-52565	US-10-282-122A-62955	US-09-809-665A-109
	Sequence 53063, A	Sequence 10312, A	Sequence 56702, A	Sequence 59492, A	Sequence 47769, A	Sequence 77935, A	Sequence 68773, A	Sequence 51291, A	Sequence 56195, A	Sequence 10989, A	Sequence 58109, A	Sequence 43, Appl	Sequence 4837, A	Sequence 10882, A	Sequence 42421, A	Sequence 50508, A	Sequence 4975, Ap	Sequence 60483, A	Sequence 171, Appl	Sequence 171, Appl	Sequence 67113, A	Sequence 49590, A	Sequence 65516, A	Sequence 54319, A	Sequence 45343, A	Sequence 65836, A	Sequence 36, Appl	Sequence 57601, A	Sequence 52565, A	Sequence 62955, A	Sequence 109, Appl	

#### ALIGNMENTS

RESULT 1  
US-09-907-907A-42  
Sequence 42, Application US/09907907A  
Publication No. US20030099660A1  
GENERAL INFORMATION:  
APPLICANT: Fisher, Paul B.  
TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A  
FILE REFERENCE: A34584-A-PCT-USA (070050.1664)  
CURRENT FILING DATE: 2001-07-16  
PRIORITY APPLICATION NUMBER: US 09/907,907A  
PRIOR FILING DATE: 1999-02-02  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 42  
LENGTH: 705  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-907-907A-42

Query Match	100.0%	Score 3557	DB 10	Length 705
Best Local Similarity	100.0%	Pred. No. 7.8e-270		
Matches	705	Conservative 0	Mismatches 0	Indels 0

Query 1 DGPFLPRRDRALTOLOVRAIMSSAGRAVAVDGNRLTEISSGKLAFAGSAVVOGSD 60  
Db 1 DGPFLPRRDRALTOLOVRAIMSSAGRAVAVDGNRLTEISSGKLAFAGSAVVOGSD 60  
Query TAVVTAWSKTPSPSPQMPVAVYRQKAAAGRIPTNYLAREVGTSDKELTSTRIDRS 120  
Db 61 TAVVTAWSKTPSPSPQMPVAVYRQKAAAGRIPTNYLAREVGTSDKELTSTRIDRS 120  
Query IRPLPAGFYDVOVLCNLAVDGVBEDVLAINGASVALSLSDIPMNGVGAVRIGIID 180  
Db 121 IRPLPAGFYDVOVLCNLAVDGVBEDVLAINGASVALSLSDIPMNGVGAVRIGIID 180

Db 121 IRPLFPAGYFEDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNPGVGAIRIGIID 180  
Qy 181 GEYVNPTRKEMSSSTLNLVAGAPKSOIWMLEASAEINILQODFCHAIKGVKTYTOOIIQ 240  
Db 181 GEYVNPTRKEMSSSTLNLVAGAPKSOIWMLEASAEINILQODFCHAIKGVKTYTOOIIQ 240  
Qy 241 GIQQLVETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300  
Db 241 GIQQLVETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300  
Qy 301 LDTEBOLKEKPEADPEYIIIESFNVAKEVRSIVLNEYKCDGRDLTSLRNVSCEYDMF 360  
Db 301 LDTEBOLKEKPEADPEYIIIESFNVAKEVRSIVLNEYKCDGRDLTSLRNVSCEYDMF 360  
Qy 361 KTLHGSALFQRGQOVLCCTVTFDLSBSGIKSDQVITAINGIKDNFMHYEFPYATNEI 420  
Db 361 KTLHGSALFQRGQOVLCCTVTFDLSBSGIKSDQVITAINGIKDNFMHYEFPYATNEI 420  
Qy 421 GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLESNGSSMASACGSLALMD 480  
Db 421 GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLESNGSSMASACGSLALMD 480  
Qy 481 SGVPISSAVAGVAILGYTKTDPEKGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540  
Db 481 SGVPISSAVAGVAILGYTKTDPEKGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540  
Qy 541 QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600  
Db 541 QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600  
Qy 601 AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660  
Db 601 AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660  
Qy 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705  
Db 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705

RESULT 2  
US-09-907a-44  
; Sequence 44, Application US/09907907A  
; Publication No. US20030099660A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Paul B.  
; APPLICANT: Leszczyniecka, Magdalena  
; TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESENCE A  
; TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF  
; FILE REFERENCE: A34584-A-PC-T-USA (070050.1664)  
; CURRENT APPLICATION NUMBER: US/09/907, 907A  
; PRIOR FILING DATE: 2001-07-16  
; PRIOR APPLICATION NUMBER: US 09/243,277  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 705  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-907-907a-44

Query Match 100.0%; Score 3557; DB 10; Length 705;  
Best Local Similarity 100.0%; Pred. No. 7.8e-270;  
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGPELPRRRDALTOLOVRALMSSAGSRAVAVDIGNRKLETSISGKLARFADGSAVVOGSD 60  
Db 1 DGPELPRRRDALTOLOVRALMSSAGSRAVAVDIGNRKLETSISGKLARFADGSAVVOGSD 60  
Qy 61 TAVVVTAVSKTSPSPQFMPVVDYRQKAAAGRIPTVYLAREVGTSDKEILTTSRIIDRS 120  
Db 61 TAVVVTAVSKTSPSPQFMPVVDYRQKAAAGRIPTVYLAREVGTSDKEILTTSRIIDRS 120

Qy 121 IRPLFPAGYFEDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNPGVGAIRIGIID 180  
Db 121 IRPLFPAGYFEDYQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPNNPGVGAIRIGIID 180  
Qy 181 GEYVNPTRKEMSSSTLNLVAGAPKSOIWMLEASAEINILQODFCHAIKGVKTYTOOIIQ 240  
Db 181 GEYVNPTRKEMSSSTLNLVAGAPKSOIWMLEASAEINILQODFCHAIKGVKTYTOOIIQ 240  
Qy 241 GIQQLVETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300  
Db 241 GIQQLVETGVTKTPQKLTFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300  
Qy 301 LDTEBOLKEKPEADPEYIIIESFNVAKEVRSIVLNEYKCDGRDLTSLRNVSCEYDMF 360  
Db 301 LDTEBOLKEKPEADPEYIIIESFNVAKEVRSIVLNEYKCDGRDLTSLRNVSCEYDMF 360  
Qy 361 KTLHGSALFQRGQOVLCCTVTFDLSBSGIKSDQVITAINGIKDNFMHYEFPYATNEI 420  
Db 361 KTLHGSALFQRGQOVLCCTVTFDLSBSGIKSDQVITAINGIKDNFMHYEFPYATNEI 420  
Qy 421 GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLESNGSSMASACGSLALMD 480  
Db 421 GKVTGLNRRELGHGALAEKALYPVIRPDPTIRVTSEVLESNGSSMASACGSLALMD 480  
Qy 481 SGVPISSAVAGVAILGYTKTDPEKGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540  
Db 481 SGVPISSAVAGVAILGYTKTDPEKGEIEDYRLTDIIGIEDYNGDMPFKIAGTKKGTAL 540  
Qy 541 QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600  
Db 541 QADIKLPGIPIKIYMEAIQOASVAKKEILOIMNKTISKPRASRENGPVEVTVQVPLSKR 600  
Qy 601 AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660  
Db 601 AKFVPGGYNLKKLOAETGVITISQVDETFSVFAPTPSVHMEADFTIEICKDOEOOLE 660  
Qy 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705  
Db 661 FGAVYTTATITEIRDTGVWVKLYPMNTAVLLNHTQLDNERLNIILP 705

RESULT 3  
US-10-408-765A-1791  
; Sequence 1791, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Faby, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Wainock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408, 765A  
; PRIOR FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1791  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-1791

Query Match 64.2%; Score 2285; DB 16; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2.8e-170;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGPELPRRRDALTOLOVRALMSSAGSRAVAVDIGNRKLETSISGKLARFADGSAVVOGSD 60  
Db 18 DGPELPRRRDALTOLOVRALMSSAGSRAVAVDIGNRKLETSISGKLARFADGSAVVOGSD 77

```

Oy 61 TAVVTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTYLRREVGTSDEKILTSRIIDS 120
Db 78 TAVVTAVSKTSPSPQFMPVLVDYRQKAAAGRIPTYLRREVGTSDEKILTSRIIDS 137
Oy 121 IRPLPAGYFYDTQVLCNLAAVDGVEBPDLAINGASVALSLSDIPMNGPVGAVRIGIID 180
Db 138 IRPLPAGYFYDTQVLCNLAAVDGVEBPDLAINGASVALSLSDIPMNGPVGAVRIGIID 197
Oy 181 GEYVNPTRKMSSTLTNLVVAAGAPKSGIWMLEASAEIILLODPFHALKVGKTYQOIIO 240
Db 198 GEYVNPTRKMSSTLTNLVVAAGAPKSGIWMLEASAEIILLODPFHALKVGKTYQOIIO 257
Oy 241 GIQOLVETGYTKTPQKLTFTSPPEIVYTKHLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
Db 258 GIQOLVETGYTKTPQKLTFTSPPEIVYTKHLAMERLYAVFTDYEHDKVSRDEAVNKIR 317
Oy 301 LDTEEQLEKEKPEADPEYIIESFNVAKEVFRSIVLNEYKSCDGRDLTSLRNVSCVDMMF 360
Db 318 LDTEEQLEKEKPEADPEYIIESFNVAKEVFRSIVLNEYKSCDGRDLTSLRNVSCVDMMF 377
Oy 361 KTLHGSALEFORGQOVLTCTVTPDSLESGIKSDOVITAINGIKDKFMHLHYEPPIYATNEI 420
Db 378 KTLHGSALEFORGQOVLTCTVTPDSLESGIKSDOVITAINGIKDKFMHLHYEPPIYATNEI 437
Oy 421 GKVTGLNRRELGHGALAEKALYPVLPDP 450
Db 438 GKVTGLNRRELGHGALAEKALYPVLPDP 467

```

## RESULT 4

```

US-10-437-963-117217
; Sequence 117217, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 117217
; LENGTH: 1526
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(1526)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20643C.1.dep
US-10-437-963-117217

```

```

Query March 35.6%; Score 1267; DB 16; Length 1526;
Best Local Similarity 38.4%; Pred. No. 1,1e-89;
Matches 275; Conservative 137; Mismatches 240; Indels 64; Gaps 11;
Oy 33 DIAGNRKLEISSGKLARFADGSAVVGSDTAVVMT-AVSKTSPSPQFMPVLVDYRQKAAA 91
Db 66 EIGGRVISTFETGKARFANGSVISMDTHVLSLTAASKSBRVDFPLVTDVGEKQYA 125
Oy 92 AGRIPNTYLRREVTSDEKILTSRIIDSIRPLPAGYFYDTQVLCNLAAVDGVEBPVL 151
Db 126 QGVIPPTVYMRREGAPKEREELLCGRILIDRIRPLPAGYFYHEQVAVNVISDGDQDPVLM 185

```

```

Oy 152 AINGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKMSSTLTNLVVAAGAPKSGIWM 211
Db 186 AANSSAALMLSDIPMNGPIGVIRGRIDGNFVLPNTVDGLSDNLVVA-CSRDKTLM 244
Oy 212 LEASAEIILLODPFHALKVGKTYQOIIOGIQOLVETGYTKTPQKLTFTSPPEIVYTKH 271
Db 245 IDVQARELTERDLQAGMKLAFAEAVKICINPOLRLAKRAG-KKKKEXYKISLSDSXEYKIR 303
Oy 272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEBOLEKEKPEADPEYIIESFNVAKEV 331
Db 304 TLSAPLEEVFTDSTYGKFERGELBENTTOSVKKALESCDEDSKFLAKAVDTRKQVI 363
Oy 332 RSVLNEYKSCDGRDLTSLRNVSCVDMMFHTLHGSALEFORGQOVLTCTVTPDSLESGIKS 391
Db 364 RKRILIEKGLRVDGRLDEVRPLCYESSYPIILHGSALEFORGQOVLTCTVTL-----GAPG 418
Oy 392 D-QVITAINGIKDKFMHLHYEPPIYATNEIETKVTGLNRRELGH----- 433
Db 419 DAQRILDSIVGPPTGRFMLHVSFPFESINEVAKRGLNRREVGHVYRNKPIRLGDELRYWI 478
Oy 434 -----GALAEKALYPVLP--RDPPTIRVTSVYLESNGS 465
Db 479 EHKALDPTGELLNRLYFWERQWLSAGTLAEKALLAVLPBEGEPPTVRAVSEVMAADGS 538
Oy 466 SSMAACGSGSLALNDGVPISSAVAGVALGLVTKTDEPEKGEIEDYRLTLIDIEDYNGD 525
Db 539 TSMAVCGSGMALNDAGIPVAEHAAGVSVGLVSEVDQGTGDISSYRITDILGIEDHGD 598
Oy 526 MFKIAGTNKGIYALQADIKLPGIPIKIWEAIOQASVAKKEIIOINKTKISKRPAKKE 585
Db 599 MFKIAGTRRGITAIQIDIKPAGIPLDICELEPARKARQIILDRMOQESSARAFAFDG 658
Oy 586 NGPVETVQVPLSRKAFVVGCGYNLKKLOAEFTGVTSIQVDEFTSVAPRPSVHGEARD 645
Db 659 SSPRLATISFSSDSLRLKL----FHRKKIIBETGARRS-VSDGVTYAKQPIPTDKALE 713
Oy 646 FITEICKDQEQLEFGAVVYATITTEIRDTGVWVKLYPNMTAVLLHNTQDNERLN 701
Db 714 KVEFLV----GRELEVGRTYKGVVSIKEYGAFVBFNGCGG-LHHSLSHDKVS 764

```

## RESULT 5

```

US-10-282-122A-51965
; Sequence 51965, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Hiansu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 51965
; LENGTH: 703
; TYPE: PR1
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51965

Query Match      33.9%; Score 1204.5; DB 15; Length 703;
Best Local Similarity 39.6%; Pred. No. 2.6e-85;
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

Qy 33 DLGRKLEISSGKLARFADGSAVVGSGPTAVMTAVSTKSPSPS-QEMLVVDYRQKAA 91
Db 8 DIARCKLVKVECKGTGMLSNCFISYDITVVMVNVNASEKREGIDFPLSIEBEROYS 67

Qy 92 AGRIPYVLRREVSTDEKELTSRIIDRSIRPLEPAGYFYDTQVLCNLAVDGVNEPDVL 151
Db 68 VGRIPGVKRGKRGSEKSTIHARADIRPLRPFKGYRNDVQVCTVMSVEQDNLEPIL 127

Qy 152 AINASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVAGAPKSQIVM 211
Db 128 AMNAGSVALCLSDIPFTPTVAVTAVSGCIDGKFLNPLTEBERKSSLDLTVG-ATNERVYM 186

Qy 212 LEASAEENLQDDFCHAIKVGKVTQOIIQGIQOLVKEGVTKRTPOKLFTSPSEIYKTH 271
Db 187 LEAGADEIPEDMLAALIDFGNACODIVAFQEKMKKEGCKKATPE-LYHKEEIEKQVT 245

Qy 272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTBEOLKEKPEADPYEIIESFNVAKEVF 331
Db 246 EFAPESIKEIM-YITDRDENLRLREIKESINBPKEKYPD-DGADIDEVVTYLOKKVV 302

Qy 332 RSIYLVNEYKCDGDLISLRVNSCEVDMFKTLHGSALFQRCQTOVLCTVTDSLESIGKS 391
Db 303 RNMLLKEHRBDGRFDEIRPISCDVDLPRTHSGGLFTRGLTOVMTVTLV---GPIGD 358

Qy 392 DQVTAINGIKDKNPMLEHYEPFPAATNEIGKVTGAINRELGHALAELKALPVIP-RDF 449
Db 359 AQVIDGLGVESKRYMHYINFPYISGEVKEPLRGPNRKEIGHALAEKALVPLIPSEEEF 418

Qy 450 PFTIRVISEVLESNGSSSMASACGSLALMDSGVPISAVAGVALIGLVTKTDPKGEIED 509
Db 419 PYTIRVISEVLESNGSSSQASVCGSTLALMDAGVPIKRPAGIAMGLTSSDLKEAV-- 476

Qy 510 YRLITLDIGIDVNDMDFKIAGTNKGITTAQADIKLPGIPIKIVMEAIQASVAKKEIL 569
Db 477 ---ITDIOGLDPRGDMDFKVAEGTEKGITTAQVDTKIHGSLSKYCIKTAINDARARLFI 533

Qy 570 QIMNKITSKPRASKENGVVETVOVPLSKRAKFPVPGCVNLKTLQATGVTISQVDEET 629
Db 534 EKMAACINPEPKKEISTYAPRAYTINIDDKRTILIGTGKTIINKIIEETGKIIIREDET 593

Qy 630 FSVFAPSPVMEHARDFTEICDKDOEQLEFGAVYATATTEIRDTGVNVLVPMNTAVL 689
Db 594 VFVLSSDASDANRALKMIIDLTKD----VAVGEVYIGKVTKINPFGAFVEVLGKGS-L 647

Qy 690 LAHTQOLDNERLN 701
Db 648 VHSIKLDINKVN 659

RESULT 6
US-09-815-242-5443
```

```

; Sequence 5443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5443
; LENGTH: 696
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-5443

Query Match      33.5%; Score 1191.5; DB 9; Length 696;
Best Local Similarity 39.2%; Pred. No. 2.7e-84;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

Qy 24 SAGRAVAVDLGNRKLEISSGKLARFADGSAVVGSGPTAVMTAVSTKSPSPSPFMLV 83
Db 2 SOEKRYVKTWAGSLTIEYQOLAKQANGAVLVRYGTVVLSATATASKERPDGDFPLTV 61

Qy 84 DYRQKAAAGRIPNYVLRREVSTDEKELTSRIIDRSIRPLEPAGYFYDTQVLCNLAVD 143
Db 62 NYEKMAYAGKIRGGFKRGRGDDATYTLARLIDRIFRFLFKGYHDVQIMMVLISAD 121

Qy 144 GVNEPDVLAINGASVALSLSDIPMNGPVGAVRIGIIDGEYVNPTRKEMSSSTLNLVAG 203
Db 122 PDGSPQAAAMIGSSMALSVSDIPQGPVAGVNGYIDGKTYIINPTVEKEKVSRLDEAVG 181

Qy 204 APLSQIVMLEASAEENLQDDFCHAIKVG-----VYITQOIIQGIQOLVKEGVTKRTP 256
Db 182 -HDAVAVMVEAGASEITEOELBAIFPGHEIORLVDFOQOIVDHIQPVKOE----- 232

Qy 257 QKLETPS-----PEIVKTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTBEOLKE 309
Db 233 ---FIPARDEBALVERIKSLTEKGLKEVYLF-----DIQQDNENLDNK---EELVNE 281

Qy 310 KFEPAE-----YEIIESFNVAKEVFRSIVLNEYKRCDDGLTSLRNVSCVDMKTL 363
Db 282 FIDBEDENELRLKEVYALINELVKEVRLIADKIRPDGRKDEIRPLDSEVGLPRT 341

Qy 364 HGSALFQRCQTOVLCTVTFSLSGSIKSDQYITAINKIKKNPMLEHYEPFPAATNEIGKV 423
Db 342 HGSGLFTRGQTOALSVTLTAL-----GDYQIIDLGLGPEBEKRFMHYINFPNFSVGETGPV 397

Qy 424 TGNRRELGHALAELKALYVIP--RDEPFTIRVTSVLESNGSSSMASACGSLALMDS 481
Db 398 RAGRREIGHALGERALKTYIIPDTADFPYTIIVSEVLESNGSSSQASISGSLALMDA 457
```



```

Oy 482 GVPISAAVAIGAVTATKDPKGE1EDYRLTLDIGIDVNGDMPFKLAGTNKGITLQ 541
Db 458 GVPICAPAGIANGMLVTEH-----STLTLDIGMDALGDMDPKFAGTKEGTLAQ 510
Oy 542 ADIKLPGIPIKIWEALIQASVAKKEIIQINMKTISKPRASKEKNGPVVEYQVPLSKRA 601
Db 511 MDIKIDGLTREIIEEALQARGRLEINMHMLQITIDQPTLESAVPKVVMTIKDKIR 570
Oy 602 KFGVGGGYNLKLQALFQGTISQVDEFTSVAPRPPSVNHEARDFTIELCKDQEQLEF 661
Db 571 DVIGGGKKINEI IDETGVKLIDEDGTLFICAVQAMNRRREIIEITR-----EAEV 625
Oy 662 GAVVYATITREIRDTGVMYVLYPNMTAVLLHNTQLDNERL 700
Db 626 GQTYQATVKRIEKGAFGLFPKQDA-LIHISQISIKRI 663

RESULT 7
US-09-815-242-12325
; Sequence 12325, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl U.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12325
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12325

; Query Match 33.5%; Score 1191.5; DB 9; Length 698;
; Best Local Similarity 39.2%; Pred. No. 2.7e-84;
; Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12

Oy 24 SAGSRAVAVDIGNRLERISGGLAFAPDSAVVQSGDPTVWMTAVASKTRPSSQPMPLV 83
Db 2 SQEKKVFTENAGRSLLTETQGLAQANCAVLRIGDTVYVLTATASKEPRGDDPFLV 61

Oy 84 DYKQKAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLAVD 143
Db 62 NYEEKVAAGKPGCFKKRGRRPGDATTALIRPIRPLPKGYKHQVQIMNMYLSAD 121

Oy 144 GVNEEDVLAINGASVALSLSDIPMNGPVCARVIGIIDGVYVNPVRKEMSSSTLNLVAVG 203
Db 122 PDCSQMAMTIGSSVALSVSDIPFGQPIAGVAVGYIDGYIINPVEEKVSRLLDEAVG 181

```

```

QY      204  APKSIVWLESAENIILQODFCIAIKVG-----VKYQQQIIQGIQOLVKEGVTKRTF 256
Db      182  -HKAVMNVKASITSDTEBMLEAFPGHEEIQRLVDFQOQIVDHQIPAKOE----- 232
QY      257  OKLFTPS-----PEIVKYTHKLMERLYAVFTDYEHDKVSRDEAVNKIRLDTESQLE 309
Db      233  ---FIPARDEALVERIKSLTEBEKLIKETVLTF-----DKQORDENMLDMIK---EEIVNE 281
QY      310  KEPEADP-----VEIIESFNVAKEVPRSIYLVNEKYKRCDDGRDLSLRNVSCGEVDFKTL 363
Db      282  FIDEDDPENNELLIKVVAILNELVKEEVRRLLADEKIRPDGRKRPDEIRLPDSEVGLPRT 341
QY      364  HGSALFQGGQOVLCCTVTFPDSLESIGKSDQVITAINGIIDKXFMFLYEFEPYATNEIGKV 423
Db      342  HOSGLFTFGQOVALSVLTLGAL-----GDYQLDGLGLOPEBKRFMHNYHNPNSVGETGPV 397
QY      424  TGLNRELIGHGALAEKALYVPVIR-RDPEFTIRVTSVELESNGSSSMASACGSLALMDS 481
Db      398  RAPGRREIRIGHGALGRALKYIIPDPADEPPTIRIVSEVLESNGSSSOASICGSTLLAMD 457
QY      482  GVPISSAAGVAILGIVTTPDKPEKGEIEDYRLLTDILIGIEDNVGDMDFKLAGTNKGTALQ 541
Db      458  GVPIKAPAGIAMLGVLTRED-----STYLLTDIQGMDALGDMDFKAGTKEGTALQ 510
QY      542  ADIKLPGIPKIVMAIQOASVAKKEIILQNMKTISKPPASRKENGPVVEYQVPSKDA 601
Db      511  MDKIDGTRTEIRIEALLDQARRGRLEINMHMLQITDQPTBELSAYAPKVYTMIRKDKIR 570
QY      602  KFGVPGYNLKKLQAEFTGVTTISQVDEETFSVAPFPSPVMHEARDPTEIKCKDQEOQLF 661
Db      571  DVIGGPGKKINEIIDEYGVKLDIEQDGTFIFIGAVQAMINRAREIIEETR-----EAEV 625
QY      662  GAVYATITTEIRDGSVMKLYPNMTAVLLHNTQDNRL 700
Db      626  GQTYQATVKRIEYKGFVGLFPGKDA-LHHSIQISKNR 663

RESULT 8
US-10-282-122A-44413
; Sequence 44413, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foretsh, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44413
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-44413

Query Match      33.5%; Score 1190.5; DB 15; Length 698;
Best Local Similarity 39.1%; Pred. No. 3,3e-84;
Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;

QY 24 SAGRAVAVDLGNKRLKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQFMPVLV 83
DB 2 SOEKKVFETWAGRSLETFEQLAQKANGAVLVRYGDTVLSTATASKERPDGDFPLTV 61
QY 84 DYQKAAAAGRIPTNYLRREVGTSDEKELTSRIIDRSIRLPFPAGYFVDQVLCNLAVD 143
DB 62 NYEKMTAAGKIPGCFKRRGPRGDDATLTARLIDRPIRPLFPKGYNHVDQIMNVLSAD 121
QY 144 GVNEPDVLAINGASVALSLSDIPMNGFVAVRIGIIDGEYVNVPTREKSSSTLNLVAG 203
DB 122 PDGSPMAAMIGSMALSVSDIPFGPIAGVNVYIDGKVIINTVEKEKSRDLDEAVG 181
QY 204 APSQIVMLEASBNIIQDFCHAIKVG-----VKYTOQIIQGIQQLVKEGVTKRTP 256
DB 182 -HKDAVMVWAGASBETESMLEAIFGHEIRLVAFOQEIIDHIOPVKOE----- 232
QY 257 QKLFPS-----PEIVKYTHKLAMERLAVFTDYEHDKSRDEAVNKIRLDEQOLKE 309
DB 233 ---FIPARDEALVERKYSLTEEKGKLETVLTF---DKQORDENLNLK---BEIVNE 281
QY 310 KFEPADP-----YEIIESFNVAKEVERSIVLNEYKRCDSRDLTSLRNVSCVDMFKTL 363
DB 282 FIDEDEPENELLIEGVVAIINELVKEVRRLADEKIRPDGRKDEIRPLDSEVGIPLRT 341
QY 364 HGSALFQRGOTVLCVTVPDSLESIGKSDQVTAINGIKDNFMILHYEPPYATNEIGKV 423
DB 342 HGSALFQRGOTVLCVTVPDSLESIGKSDQVTAINGIKDNFMILHYEPPYATNEIGKV 397
QY 424 TGLNRRELGHGALAEKALYPIF--RDPPFIRYTSVLENSGSSSMASACGSLAMD 481
DB 398 RAPGRREIGHGALGERLAKYIIPPTADFPYTRIVSEVLESNGSSSQASICGSLTALMDA 457
QY 482 GVPISAVAGVAGIYVTKTDEPEKEIEDYRLTDLIIGIEDYNGMDFKIAGTKNGITALQ 541
DB 458 GVPISAVAGVAGIYVTKTDEPEPEKEIEDYRLTDLIIGIEDYNGMDFKIAGTKNGITALQ 510
QY 542 ADILPGLPIKIVMEAIQOASVAKKEIIQIINMKTISKPRASRKENGVPVETVQVPLSKRA 601
DB 511 MDIKIDGLTREIIEEALQARGRLIEINMHLQITIDPRTLSAAPVAVMTIKPKIR 570
QY 602 KFPVPGGNLKKLOAEFGVTTISQVDETFVSFAFPTSPVMEHARDFITIEICDDDEQOLEF 661
DB 571 DVIGPGGKKNIEIIDETGVKLDIFQDGTIFGAVDQAMINARBIETIR-----EAEV 625
QY 662 GAVVTATITEIRDTGVWVKLYPNNMTAVLNLHTQDLNRL 700
DB 626 GQTYQATYKRIEKYGAFVGLFPGKDA-LIHISQISKRI 663

RESULT 9
US-10-282-122A-70948
; Sequence 70948, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70948
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70948

Query Match      32.8%; Score 1165.5; DB 15; Length 701;
Best Local Similarity 38.7%; Pred. No. 3e-82;
Matches 270; Conservative 123; Mismatches 250; Indels 55; Gaps 13;

QY 24 SAGRAVAVDLGNKRLKLEISSGKLARFADGSAVVGSDTAVMTAVSKTKSPSQFMPVLV 83
DB 2 SOEKKVFETWAGRSLETFEQLAQKANGAVLVRYGDTVLSTATASKERPDGDFPLTV 61
QY 84 DYQKAAAAGRIPTNYLRREVGTSDEKELTSRIIDRSIRLPFPAGYFVDQVLCNLAVD 143
DB 62 NYEKMTAAGKIPGCFKRRGPRGDDATLTARLIDRPIRPLFPKGYNHVDQIMNVLSAD 121
QY 144 GVNEPDVLAINGASVALSLSDIPMNGFVAVRIGIIDGEYVNVPTREKSSSTLNLVAG 203
DB 122 PDGSPMAAMIGSMALSVSDIPFGPIAGVNVYIDGKVIINTVEKEKSRDLDEAVG 181
QY 204 APSQIVMLEASBNIIQDFCHAIKVG-----VKYTOQIIQGIQQLVKEGVTKRTP 256
DB 182 -HKDAVMVWAGASBETESMLEAIFGHEIRLVAFOQEIIDHIOPVKOE----- 232
QY 257 QKLFPS-----PEIVKYTHKLAMER--LYAVFTDYEHDKSRDEAVNKIRLDEQOLKEF 311
DB 233 ---FVPERDEDLVEKYKSLTEDEKGLDVTLT---FDQORDENLNLK---BEVGHFL 283
QY 312 PADP-----YEIIESFNVAKEVERSIVLNEYKRCDSRDLTSLRNVSCVDMFKTLAG 365
DB 284 DEEDPENETLVKGVVAILNDLIEKVRRLADEKIRPDGRKNDVEIRLSEVGLLPRAHG 343
```

```

QY 366 SALFORQOTVLTCTVETDSSESIGKSDQVITAINIGIKDKNFMLHYEPPVATNEIGKVTG 425
D 344 SGLFTRGQTALSLTLTGAL----GDYQLIDGLPBEVEKRMHMYNPNFVSGETGPVRA 399
QY 426 LNRRLGHGALAEALYVLP--RDPPFTIRVTEVLESNGSSSMASACGSLALMSGV 483
D 400 PGRREIGHGALGERALYIIPDTODFPYTRIVSEVLESNGSSSQASICGSLTALMDAGV 459
QY 484 PISAVAGVAVIGLTKTDPKEGEIEDRYLLDIDIGIDYNDMDPKIAGTNKGTALOAD 543
D 460 PIRAPVAGVIGLTKTDP-----STTLIDIGMEDALGDMDFKVAIGTKDGTITAIOMD 512
QY 544 IKLPGPIKIVMEAIQOASVAKKEILOIMNKTISKPRARENGPVVETVQVPLSKDAKF 603
D 513 IKIDGLREVIEBALAQOGLALIMDMHTIEQPREELSAVAPKVTWSINDKIRDV 572
QY 604 VEGGVNKLQOAGTGTISQVDEBETSVPAPTSVMHEARDFTETICKDOBOGLEPGA 663
D 573 IGPGRKKINEIIDEQVGLDIEODGTIFIGAVDQAMINRAKEIIEDITR-----EAEVGG 627
QY 664 VYATITIRDTGVVTKLYPMNTAVLLAHNTQIDNERLN 701
D 628 VYHAKVKRIEKYGAFFELPEKDA-LAHISQISGERIN 664

```

## RESULT 10

```

US-10-282-122A-71588
Sequence 71588, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forgyh, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 71588
LENGTH: 692
TYPE: PR
ORGANISM: Staphylococcus haemolyticus

```

```

FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (470)..(470)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-71588

```

```

Query Match 32.7%; Score 1161.5; DB 15; Length 692;
Best Local Similarity 38.5%; Pred. No. 6,1e-82;
Matches 262; Conservative 126; Mismatches 247; Indels 45; Gaps 11;

```

```

QY 37 RKLRISSGKLARPADGSAVVOGDTAVMTAVSKTKSPQCFMPLVVDYKQAAAARIP 96
D 9 RSLTIEIGOLAKQANGAVLVRGDTVVLSTAVASKEBRDDFPPLTVYEEKYMAAKIP 68
QY 97 TNYLRREVGSDKEILRSRIIDSIRPLFPAGFYDQVLCNLLAVDGNVBDVLAINGA 156
D 69 GGFGRKRGPRGDEATTLARLIDRPIRLFPKGRHNDVQIMTVLSADPDCSPFMAATIGS 128
QY 157 SVALSLSDIPWNGVGAIRIGIDGEVYVNPTRKEMSSSTLNLVAGAPRSQIVMLEASA 216
D 129 SMALSVSDIPFGIAGVNVGVYDGKVINPTEBEKESRLDELVAG-HKDAVNVMEAGA 187
QY 217 ENLIQODFCHAIKVG-----VKYTOQIIIGIQQLVKEVGTVKRTPOKLFPSPELVKY 269
D 188 SEITEKEMLEAIFFGHDEIRKLVAFOEEVVAHIQPVKKEFVPEVDEALV-----SRVKT 243
QY 270 THKLAMERLVAVFTDYHDKVSDEAVNKRIDTBEOLKEKPEADPY-----EIESF 323
D 244 TEKGLKETVLT-----DKQQRDENLDTKAEIATEFVD--EADENELLIDEVTAII 295
QY 324 NVVAKEVFRSIVNEVRCGRDLTSLRNVSCEDVMEKTLGSLFORQOTVLTCTVTFD 383
D 296 NDLVKEVRRLIADDEKIRPPGRKPRDEIRPLESEVGLIPRAHGSLFRGQTQALSVTLIG 355
QY 384 SLBSGKSDQVITAINIGIKDKNFMLHYEPPVATNEIGKVTGLNRRRLGHGALAEALY 443
D 356 AL-----GDYQLIDGLGEGEKEKRFMHYHNPFPVSGETGPVRAVGRREIGHGALGERAL 411
QY 444 VIP--RDPPFTIRVTSVLESNGSSSMASACGSLALMDSGVPISSAVAAVAGLVTKTD 501
D 412 IIPDTTEFPYTVIVSVLESNGSSSQASICGSLTALMDGVPIKAAVAAIANGLVTRXD 471
QY 502 PEKGEIEDRYLLDIDIGIDYNDMDPKIAGTNKGTITAIQADIKLPGIPKIIVMEAIQOA 561
D 472 SNPD-----LIDIGMEDALGDMDFKVAIGTKDGTITAEIABEQA 524
QY 562 SVAKKEILOIMNKTISKPRARENGPVVETVQVPLSKRAKFFVGGVNLKULOAGETGT 621
D 525 ROGRALIMDMHQLTIDDPKELSAVAPKVEIMHIKPEKIRDIVIGGGKKINEIIDEQVGL 584
QY 622 ISQVDEBETSVPAPTSVMHEARDFTETICKDOBOGLEGAAYTATITETIRDTGVNKL 681
D 585 LDIEODGTIFIGIDQDMINRAKEIIEDITR-----EAEVGYVNAKRIEKYGAFFEL 639
QY 682 YPMNTAVLLAHNTQIDNERLN 701
D 640 FAGKDA-LAHISQISGERIN 658

```

## RESULT 11

```

US-10-282-122A-75529
Sequence 75529, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant

```

```
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 75529
/ LENGTH: 711
/ TYPE: PRT
/ ORGANISM: Salmonella typhi
/ US-10-282-122A-75529

Query Match      32.6%; Score 1161; DB 15; Length 711;
Best Local Similarity 39.9%; Pred. No. 6.9e-82;
Matches 270; Conservative 118; Mismatches 249; Indels 40; Gaps 11;

QY 35 GNRKLEISSGKLAFFADGSANVOSGDTAVWVTAISKTKPSPQ-FMPLVVDYRQAKAAAG 93
DB 12 GQHTVLTLETGMARQATAAVVWSMDTAVFVTVGQKAKRGDFFPLVNVYQERTYAAAG 71
QY 94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEPDVLA1 153
DB 72 RIPSFRRRGRSREGTLLARLIDRVRPLFPBGFNVEQVATVSVNPQVNPDIYAM 131
QY 154 NGASVALSLSDIPWNGEVAVRIGIIDGEYVNVPTREKMSSTLNLVVAAPKSQIYMLE 213
DB 132 IGASALSLSGIPFNGPIGAVRVGINDQVYVLPDDELKESKLDLVVAGT-EAAVLMVE 190
QY 214 ASANIIQDQPCFAIKRGVKTQOIIQGIQOLVKEIVTKRTPQKLTPSEIYK----- 268
DB 191 SEAEILSEDTMLGAVFGHEQOQVVOIAINDLVKAGPRADWG-----PEAVNDALMA 244
QY 269 YTHGLAMERT---YAVFTDYEDHKVSRDEAVNKLRLDTEBOLKKEKFPADPEYETIESFNV 325
DB 245 RVALAASRLSDAYRI-----TDQERYAQVDVAKSETIIELDIAEDELTDANEIGETIIHA 299
QY 326 VAKEVFESIVINEYKRCGRDLTSLRVNSCEVDFMFKLHGSAIFQRCQTVQLCTVTPDSL 385
DB 300 IEKVIVSVRVLGAPRIDGREKDMIRGLDVRTGLPRTHGSAFLPRTGBTQALVATATLGT- 358
QY 386 ESGIKSQOVTAINGIDKXNMLHYEPPTATNEIGKVTGNRRELGHGALAELALPYVI 445
DB 359 ---ARDQVVDLMDGERTDSFLFHYNPFPYVSGGTGMVGSFKRREIGHGRACKGVLAVM 415
QY 446 P--RDPPTIRVTSEVNESSSSSMASACGSLALMDGVPISSAVAGVALGATKTPDPE 503
DB 416 PDMDKFPYTVAVVEITESNSSSSMASVCGSLALMDGVPIKAAVAGIAGLVKEGD-- 473
QY 504 KGEIEDYRLLTDILIGIEDYNGDMDFKLAGTNKGITLALQADIKLPGIPIKIYMEAIQOASV 563
```

```
DB 474 -----NIVVLSDTLGDBDHLGDMDFVAGSRDGLSALQNDIKLEGITKEIMQVALNDAGK 528
QY 564 AKKEIILQIMKTIKSPRASHKENGPFVETQVPLSKRAKFGVGNLKKLQAEYGTIIS 623
DB 529 ARHLILGVMQALNAPRGDISFAPRHTIKISTDKIKDVIKGGSVIRALTEETGTIE 588
QY 624 QVDEETFSVPAPFPSSVMEHARDITTEICNDQDQQLFEFCAVYATATTEIDTGVMYKLYP 683
DB 589 IEDDGTIVKIAATGGEKAKVARIIRIEI-----PABIEVGRIYNSKVYRIYDFGAFVAGG 643
QY 684 NMTAVILLNTOIDNERL 700
DB 644 KGEK-LVHISQIADKRV 659

RESULT 12
US-09-815-242-13768
Sequence 13768, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zykkind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13768
/ LENGTH: 721
/ TYPE: PRT
/ ORGANISM: Salmonella typhi
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (1)..(721)
/ OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13768

Query Match      32.6%; Score 1158; DB 9; Length 721;
Best Local Similarity 39.9%; Pred. No. 1.2e-81;
Matches 270; Conservative 117; Mismatches 250; Indels 40; Gaps 11;

QY 35 GNRKLEISSGKLAFFADGSANVOSGDTAVWVTAISKTKPSPQ-FMPLVVDYRQAKAAAG 93
DB 22 GQHTVLTLETGMARQATAAVVWSMDTAVFVTVGQKAKRGDFFPLVNVYQERTYAAAG 81
QY 94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGFYDTQVLCNLAVDGVNEPDVLA1 153
DB 82 RIPSFRRRGRSREGTLLARLIDRVRPLFPBGFNVEQVATVSVNPQVNPDIYAM 141
QY 154 NGASVALSLSDIPWNGEVAVRIGIIDGEYVNVPTREKMSSTLNLVVAAPKSQIYMLE 213
```

```

Db 142 IGASALSLISGIFPNIGIGARVGYINDQVYLANFTQBELKESKLDLVVAGT-EAAVLAMVE 200
Qy 214 ASAMENIQDPFCAIKVGVYKTOQIIIGIQOVLVETGVTKTPKLTTPSPBYK----- 268
Db 201 SEABELSEDMLGVVGHGHOQVIAINDLVYKAKPRMDW-----FEAVNDALMA 254
Qy 269 YTHKLMERL---YAVTDVENDKVSDEAVNKRLRTEBOLKPKPEADPYEIESFN 325
Db 255 RVALLAESRLSDAYRI-----TDKQERYAOVDYKSETIEBOLIADETLDAEIGELHA 309
Qy 326 VAKCEPESIVLNEYKRCGRDLTSLRVNSCEVDMFKTLHGSALFQSGQVYLCVTEPDSL 385
Db 310 IEKNVARSRLVAGPRIDGRHKMIRGLDVRTGLPRTGHSALTRBETQALVATLGT- 368
Qy 386 ESGIKSDQVITTAINGIKDKNFMHYEPPIATNEIGVYGLNRELGHALAEKALYPI 445
Db 369 ---ARDAQVLDLMGERTDFFLPHYNPFPYSGVETGVVSGPKREIGHGRLAKRGVLA 425
Qy 446 P--RDPFTIRVSEVLESNGSSSMASACGSLALMDSGVPISAVAGVAGVAGVAGV 503
Db 426 PDMKFFYTVRVSSEITESHSSSMASVCGASLALMDAGVPIKAAVAGVAGVAGV 483
Qy 504 KGEIEDYRLTLGIEDYNGDMDFKTAGTNKGTALQADIKLPGIPIKIWMENIQOASV 563
Db 484 ----NYVLSLIDGEDHIGDMDFKVASGRDGSALQMDIKIEGITEIMQVVALNQAKG 538
Qy 564 AKKEILOINNKTSKPRASKEGPPVETVQVPLSKRAKRVGCGVNLKQLQAEVGTIS 623
Db 539 ARHLILVMEQAINAPRGDISFAPRIHTIKISTDKIKDVGKSGSVIRALTEBTGTIE 598
Qy 624 QVDEETSVFAPRPSVHNEARDFITEICXDDQEQOLFEGAVYATTAIEIHDTGVMLY 683
Db 599 IEDGVYKIAATDEKAKVARIIRIEI-----TAEIVGRYNSKVRIVDPFGAFVAG 653
Qy 684 NMTAVLHNTQOLDNERL 700
Db 654 GKGG-LVHISQIADKRV 669

```

```

RESULT 13
US-10-282-122A-61234
; Sequence 61234, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: Elitza.034A
; CURRENT APPLICATION NUMBER: US/10/282, 122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191, 078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206, 848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207, 727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230, 335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230, 347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242, 578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253, 625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257, 931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267, 636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61234
; LENGTH: 729
; TYPE: PR
; ORGANISM: Legionella pneumophila
US-10-282-122A-61234

```

```

Query Match 32.5%; Score 1157; DB 15; Length 729;
Best Local Similarity 38.2%; Pred. No. 1.5e-81;
Matches 258; Conservative 134; Mismatches 244; Indels 40; Gaps 12;

```

```

Qy 35 GNRKLEISGKLARFADGSAVUSGDTAVWTVAVSKTSPS-QFMPLVVDYRQKAAAG 93
Db 11 GNRKLEISGKLARFADGSAVUSGDTAVWTVAVSKTSPS-QFMPLVVDYRQKAAAG 70
Qy 94 RIPTNYLREVGTSDEKILTSRIIDSIRPLPAGYFDYQVLCNLLAVDGNBPVLA 153
Db 71 KIPGFRKRGSRSDDELTLRLIDRIRPLPAGYFDYQVLCNLLAVDGNBPVLA 110
Qy 154 NGASVALSLSDIPWNGPVAVRIGIIDEVYVNPTEKSSSTLNVVAGAPKSOIYME 213
Db 131 IGASALSLISGIFPNIGIGARVGYINDQVYLANFTQBELKESKLDLVVAGT-EAAVLAMVE 189
Qy 214 ASAMENIQDPFCAIKVGVYKTOQIIIGIQOVLVETGVTKTPKLTTPSPBYK----- 266
Db 190 SEAOBLESIDIMRGAMLYGHEMMGVKSIHELADVQSKR--PE---WKAPRIDYVKAR 244
Qy 267 VKYTHKLMERLAVAFVDEHDKVSRDEAVNKRLRTEBOLKPKPEADPYEIESFN 326
Db 245 INVARNEVAAVLI-----KDKQRYQRLDELRBOTISLALLENDELNAVIANNGEL 299
Qy 327 AKCEPESIVLNEYKRCGRDLTSLRVNSCEVDMFKTLHGSALFQSGQVYLCVTEPDSL 386
Db 300 ERSIVNRILDSGRPRIDGRHRTVPRISIRTKLEBTHGSLCTPRTGTQALVATL---- 355
Qy 387 SGIKSD-QVITTAINGIKDKNFMHYEPPIATNEIGVYGLNRELGHALAEKALYPI 445
Db 356 -GNERDAQIUDGISGESRDFMLHYNPFPYSGVETGVVSGPKREIGHGRLAKRGVLA 414
Qy 446 P--RDPFTIRVSEVLESNGSSSMASACGSLALMDSGVPISAVAGVAGVAGVAGV 503
Db 415 PDANEPFYVLRIVSEITESHSSSMASVCGASLALMDAGVPIKAAVAGVAGVAGV 472
Qy 504 KGEIEDYRLTLGIEDYNGDMDFKTAGTNKGTALQADIKLPGIPIKIWMENIQOASV 563
Db 473 ----RAVYLDLIGEDHIGDMDFKVASGRDGSALQMDIKIEGITEIMQVVALNQAKG 527
Qy 564 AKKEILOINNKTSKPRASKEGPPVETVQVPLSKRAKRVGCGVNLKQLQAEVGTIS 623
Db 528 GRTHILGVNNALAEHRTLSQAPRIITWIKVADKIRITIGKAGATIKILISTGVSID 587
Qy 624 QVDEETSVFAPRPSVHNEARDFITEICXDDQEQOLFEGAVYATTAIEIHDTGVMLY 683
Db 588 IDSQVQLSPKRALMEKQIKALI-----ABIEVGQTYGKSKIVDPGAFINL 642
Qy 684 NMTAVLHNTQOLDNER 699
Db 643 GKGG-LVHISQIADKRV 657

```

```

RESULT 14
US-10-282-122A-53063
; Sequence 53063, Application US/10282122A

```

```
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 53063
LENGTH: 703
TYPE: PRT
ORGANISM: Clostridium difficile
US-10-282-122A-53063

Query Match      32.5%; Score 1155; DB 15; Length 703;
Best Local Similarity 38.6%; Pred. No. 2e-81;
Matches 261; Conservative 136; Mismatches 249; Indels 30; Gaps 14;
```

```
Db 305 EEVRKLIYHENVAPDNRKLEIRPINCETGMIPIRAHGSALFTFGQTVLVNATLGLAGDV 364
Qy 389 IKSDQVITTAINGIKDNFMHYEPPEPYATNEIGKVTGLNRELGAGLAKEALYVIP- 446
Db 365 QKLD---GLDBEENKRYMHYNFPAYSVGEARPSGPRRIRIGHALERALLPVLPsq 420
Qy 447 RDPFPTIRVTSYLVLENGSSMASAGGSLALNDGVPRISSANVAVAIIGVTTDPKEG 506
Db 421 EEPFYAIRLVSEVSSNGSTQASVCGSTLSLDAGVPIDWAVAGIAMGLI-RHD--GK 476
Qy 507 IEDYRLITDILIGEDYNDMDPKIAGTNKGITLALQADIKLPGIPRIKIVEAIIQASVAK 566
Db 477 V---AVISDIQGMEDHGDMDFKVAGTEYGITAIQMDIKIDGIDKELIQALQAGRI 533
Qy 567 EIIQINMKTIKSPRASKENGPEVETVQVPLSRKAFVGGYNLKLQAEFTVTISQVD 626
Db 534 HILGEMRKTIQPKPELSPYAKIVKMQINPDKIKVIGGKLTITKIIDETGVKID--I 591
Qy 627 EEPFVPAPTPSV--HEADPTIEICKDQBOQLERGANYYTATITIEIRPTGVNVLKPN 684
Db 592 EQTGEVFIQIEIDMKKQDELINNIIVEP---EVGKTYKGVKSRIMNFGAFVEILPG 646
Qy 685 MTAVLHNTQDNERL 700
Db 647 KEG-LHISHIAHERV 661

RESULT 15
US-09-815-242-10312
Sequence 10312; Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: EITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10312
LENGTH: 734
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10312

Query Match      32.5%; Score 1155; DB 9; Length 734;
Best Local Similarity 39.9%; Pred. No. 2.2e-81;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;
```

```
Qy 35 GNRKLEISSGKLARFADGSAVVOGDPAAVNTAVSKTPSPQ-FMPLVVDYDQKAAAG 93
```

Db 35 GQHTVLTLETGMMAQATAAVVMSMDTAVFVTVVGQKKAKPGODFFPLTVNOERTYAAG 94  
QY 94 RIFPNVYAREVGTSDKXELTSRIIDRSIRLPFAGYFVDTOVLGNLAVDGNEDVLAI 153  
Db 95 RIGSFPRRREGRPSEGETLRLIDRPLRPLFPGFVNEGVVIAVSVNPQVNDIVAM 154  
QY 154 NGASVALSLSDIPNNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNLVAGAPKQIWMLE 213  
Db 155 IGASVALSLSGIPNNGPVGAVRIGIIDGEVYVNPTRKEMSSSTLNLVAGAPKQIWMLE 213  
QY 214 ASANILIQODFCHAIKVGVKYYTQOIIGIQQLVKRVTKTPOKLFPPSP--EIVKYT 270  
Db 214 SEAQLISEDQMLGAVVFGHEGQQOVVIQINELVKEAG---KPRWMOPEPVEPNBALNARV 269  
QY 271 HKLAMERL--YAFPTYEHKVRDEAVNKIRLDTEBQLKEKPEADPVEIIESFNVA 327  
Db 270 AALAEARISDAVRI---TDKQERYAQVDVIKSETIATLAEDETLDENELGEILHAIE 324  
QY 328 KEVRSIVLNEYKRCDSGLTSLNVSCEVDMFKTLHGSALFORGQTOVLCTVTFDSLES 387  
Db 325 KNVRSRVLAGRPIDREKXIMIGLDVRTGVLPRTGOSALFTRGETOALVTATLGT--- 381  
QY 388 GIKSDOYITAINGIKDKNFMHLHYEPFYATNEIGKVTGLNRELGHALAEKALYVPVLP- 446  
Db 382 -ARDAQVLDEIMGERTDTFLFHYNFPYYSVGETGVGSPKREIGHGRLAKRGVLAWMPD 440  
QY 447 -RDEPFTIRVTSVLESNGSSMASACGSLALMDSGVPISSAVAGVALGLVTKTDPEKG 505  
Db 441 MDKPYTVRVVSEITNESGSSSMASVCGASIALMDAGVPIKAAVAAGIAMGLVKEGD--- 496  
QY 506 EIEDYRLTDLIGIEDVNGMDFKIAGTNKGITLQADIKLPGIPIKIVMEAIQOASVAK 565  
Db 497 ---NYVLSDILGDEBDHIGMDPKVAGSRDGSIALQMDIKIEGITKEIMOVALNOAKGAR 553  
QY 566 KEILOINAKTISKPRASRKENGPEVETVQVPLSKRAKFPVPGYNLKKQDAETGVTSQV 625  
Db 554 LHLGVMEQAINABRGDISBPAPRIHTIKINPDKIKOVIGKGGSVIRALTEETGTTEIE 613  
QY 626 DEETFSVPAPFPVMEHARDPITRICKDDOEOGLEFGAVTYATITEIRDTGVWTKLYPNM 685  
Db 614 DDGTVKIAATDEKAKAIARIEI---TAIEIVGKRVYTKVTRIVDFGAFVAGGK 668  
QY 686 TAVLHNTQLDNERL 700  
Db 669 EG-LVHISQIADKRV 682

Search completed: January 28, 2005, 19:47:36  
Job time : 150 secs



**This Page Blank (uspto)**